WESKLEY DA SILVA COTRIM

INTELIGÊNCIA ARTIFICIAL APLICADA A MODELAGEM DE PROCESSOS DA INDÚSTRIA DE ALIMENTOS

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência e Tecnologia de Alimentos, para obtenção do título de *Doctor Scientiae*.

Orientador: Luis Antônio Minim

Coorientadores: Leonardo Bonato Felix
Renata Cássia Campos
Valéria P. Rodrigues Minim

VIÇOSA - MINAS GERAIS
2021

WESKLEY DA SILVA COTRIM

INTELIGÊNCIA ARTIFICIAL APLICADA A MODELAGEM DE PROCESSOS DA INDÚSTRIA DE ALIMENTOS

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência e Tecnologia de Alimentos, para obtenção do título de *Doctor Scientiae*.

APROVADA: 05 de agosto de 2021.

Assentimento:

_____

Weskley da Silva Cotrim
Autor

_____

Luis Antônio Minim
Orientador

Aos meus pais, Gerson e Joana.
À minha esposa, Keyla.
Aos meus filhos, Maria, Amanda e Henrique.
Meu passado, presente e futuro.
Este trabalho é dedicado a vocês.

AGRADECIMENTOS

"Porque dele, e por meio dele, e para ele são todas as coisas. A ele, pois, a glória eternamente. Amém!" (Carta do Apóstolo Paulo aos Romanos, Capítulo 11, Versículo 36)

A Deus pai, Santo e Criador, a Jesus Cristo, Autor e Consumador da fé, ao Espírito Santo, Consolador, pelo amor e pelas muitas bênçãos. *"Porque Deus amou ao mundo de tal maneira que deu o seu Filho unigênito, para que todo o que nele crê não pereça, mas tenha a vida eterna."* (Evangelho de Jesus Cristo Segundo o Apóstolo João, Capítulo 3, Versículo 16)

Ao meu pai, Gerson, por sempre se orgulhar do meu trabalho e me ensinar a fazer tudo com esmero e dedicação, não apenas agora que experimento os louros do sucesso, mas também quando fui um simples carroceiro. *"Vês a um homem perito na sua obra? Perante reis será posto; não entre a plebe."* (Provérbios, Capítulo 22, Versículo 29)

A minha mãe, Joana, por me ensinar a "segurar nas mãos de Deus" nos momentos mais difíceis e seguir trabalhando. Uma hora o sol aparece no horizonte. *"...mas o justo viverá pela sua fé."* (Livro do Profeta Habacuque, Capítulo 2, Versículo 4b)

A minha amada esposa, Keyla, por estar ao meu lado sempre. Sem você eu não teria chegado tão longe. *"Mulher virtuosa, quem a achará? O seu valor muito excede o de finas joias. O coração do seu marido confia nela, e não haverá falta de ganho."* (Provérbios, Capítulo 31, Versículos 10 e 11)

Aos meus filhos, Maria Eduarda, Amanda e Henrique, por me ensinarem que *"Herança do Senhor são os filhos; o fruto do ventre, seu galardão. Como flechas nas mãos do guerreiro, assim os filhos da mocidade. Feliz o homem que enche deles a sua aljava; não será envergonhado, quando pleitear com os inimigos à porta."* (Salmos, Capítulo 127, Versículos 3 a 5)

À Universidade Federal de Viçosa, ao Programa de Pós-Graduação em Ciência e Tecnologia de Alimentos, pela oportunidade de realização deste curso.

Ao professor Luis Antônio Minim, pela orientação, confiança, apoio e ensinamentos, imprescindíveis para a realização deste trabalho.

Aos professores, membros da minha comissão orientadora, Leonardo Bonato Felix, Renata Cássia Campos e Valéria Paula Rodrigues Minim pelo auxílio, atenção, avaliação crítica e valiosas sugestões, sempre oportunas.

Aos demais professores do Programa de Pós-Graduação em Ciência e Tecnologia de Alimentos pelos ensinamentos, sugestões, críticas e apoio.

Aos amigos e irmãos da Segunda Igreja Batista em Viçosa, pelos bons momentos partilhados juntos, pois *"Oh! Como é bom e agradável viverem unidos os irmãos!"* (Livro dos Salmos, Capítulo 133, Versículo 1)

Ao pastor, teólogo, mestre e amigo Luciano, e sua esposa Juliane, pela paciência e disposição para ensinar e cuidar, pois *"...ele mesmo concedeu uns para apóstolos, outros para profetas, outros para evangelistas e outros para pastores e mestres, com vistas ao aperfeiçoamento dos santos para desempenho do seu serviço, para edificação do corpo de Cristo, até que todos cheguemos à unidade da fé e do pleno conhecimento do Filho de Deus, à estatura da plenitude de Cristo..."* (Carta do Apóstolo Paulo aos Efésios, Capítulo 4, Versículos 11 a 13)

Aos colegas de laboratório, Ana, Andrea, Diogo, Jamille, Kátia, Lucidarce, Nathan, Raquel e Richard, pela constante ajuda e agradável convivência.

A todos que, de alguma forma, contribuíram para a realização deste trabalho.

"Porque foi subindo como renovo perante ele e como raiz de uma terra seca; não tinha aparência nem formosura; olhamo-lo, mas nenhuma beleza havia que nos agradasse. Era desprezado e o mais rejeitado entre os homens; homem de dores e que sabe o que é padecer; e, como um de quem os homens escondem o rosto, era desprezado, e dele não fizemos caso.

Certamente, ele tomou sobre si as nossas enfermidades e as nossas dores levou sobre si; e nós o reputávamos por aflito, ferido de Deus e oprimido. Mas ele foi traspassado pelas nossas transgressões e moído pelas nossas iniquidades; o castigo que nos traz a paz estava sobre ele, e pelas suas pisaduras fomos sarados. Todos nós andávamos desgarrados como ovelhas; cada um se desviava pelo seu caminho, mas o Senhor fez cair sobre ele a iniquidade de nós todos. Ele foi oprimido e humilhado, mas não abriu a boca; como cordeiro foi levado ao matadouro; e, como ovelha muda perante os seus tosquiadores, ele não abriu a boca. Por juízo opressor foi arrebatado, e de sua linhagem, quem dela cogitou? Porquanto foi cortado da terra dos viventes; por causa da transgressão do meu povo, foi ferido. Designaram-lhe a sepultura com os perversos, mas com o rico esteve na sua morte, posto que nunca fez injustiça, nem dolo algum se achou em sua boca.

Todavia, ao Senhor agradou moê-lo, fazendo-o enfermar; quando der ele a sua alma como oferta pelo pecado, verá a sua posteridade e prolongará os seus dias; e a vontade do Senhor prosperará nas suas mãos. Ele verá o fruto do penoso trabalho de sua alma e ficará satisfeito; o meu servo, o Justo, com o seu conhecimento, justificará a muitos, porque as iniquidades deles levará sobre si. Por isso, eu lhe darei muitos como sua parte, e com os poderosos repartirá ele o despojo, porquanto derramou a sua alma na morte; foi contado com os transgressores; contudo, levou sobre si o pecado de muitos e pelos transgressores intercedeu." (Livro do Profeta Isaías, Capítulo 53, Versículos 2 a 12)

"No dia seguinte, viu João a Jesus, que vinha para ele, e disse: Eis o Cordeiro de Deus, que tira o pecado do mundo!" (Evangelho de Jesus Cristo Segundo o Apóstolo João, Capítulo 1, Versículo 29)

"Tende em vós o mesmo sentimento que houve também em Cristo Jesus, pois ele, subsistindo em forma de Deus, não julgou como usurpação o ser igual a Deus; antes, a si mesmo se esvaziou, assumindo a forma de servo, tornando-se em semelhança de homens; e, reconhecido em figura humana, a si mesmo se humilhou, tornando-se obediente até à morte e morte de cruz. Pelo que também Deus o exaltou sobremaneira e lhe deu o nome que está acima de todo nome, para que ao nome de Jesus se dobre todo joelho, nos céus, na terra e debaixo da terra, e toda língua confesse que Jesus Cristo é Senhor, para glória de Deus Pai." (Carta do Apóstolo Paulo aos Filipenses, Capítulo 2, Versículos 5 a 11)

"Grandes e admiráveis são as tuas obras, Senhor Deus, Todo-Poderoso!
Justos e verdadeiros são os teus caminhos, ó Rei das nações!
Quem não temerá e não glorificará o teu nome, ó Senhor?
Pois só tu és santo; por isso, todas as nações virão e adorarão diante de ti, porque os teus atos de justiça se fizeram manifestos."
(Apocalipse do Apóstolo João, Capítulo 15, Versículos 3 e 4)

BIOGRAFIA

WESKLEY DA SILVA COTRIM, filho de Gerson Pinheiro de Araújo Cotrim e Joana Maria da Silva Cotrim, nasceu em Itanhém, Estado da Bahia, em 09 de fevereiro de 1979. Em março de 2001, ingressou na Universidade Federal de Viçosa – UFV no curso de Ciência e Tecnologia de Laticínios. No ano seguinte transferiu-se para o curso de Engenharia de Alimentos, vindo a concluí-lo em março de 2007. Em abril de 2007 iniciou curso de Mestrado em Ciência e Tecnologia de Alimentos na UFV, o qual foi concluído em Julho de 2010. Em agosto de 2008 iniciou sua carreira profissional como Analista de Serviços Tecnológicos em Alimentos no Centro de Tecnologia SENAI – Alimentos e Bebidas (CTS Alimentos e Bebidas), na cidade de Vassouras, RJ, onde permaneceu até julho de 2010. Em Julho de 2010 iniciou a carreira de professor do magistério superior nas Faculdades Associadas de Uberaba – FAZU, na cidade de Uberaba, MG, onde permaneceu até agosto de 2012, quando foi aprovado em concurso público para o cargo de Professor do Ensina Básico, Técnico e Tecnológico, no Instituto Federal Goiano – IFGoiano, para atuar nos cursos da área de alimentos e bebidas, na cidade de Iporá, GO. Em setembro de 2013, após aprovação em concurso público de provas e títulos para o cargo de Professor do Magistério Superior, transferiu-se para a Universidade Federal de Mato Grosso, onde atua até a presente data como professor do curso de Engenharia de Alimentos, na cidade de Barra do Garças, MT. Em julho de 2017 iniciou o curso de doutorada em Ciência e Tecnologia de Alimentos, na Universidade Federal de Viçosa – UFV, submetendo-se a defesa pública de tese no dia 05 de agosto de 2021.

RESUMO

COTRIM, Weskley da Silva, D.Sc., Universidade Federal de Viçosa, agosto de 2021. **Inteligência artificial aplicada a modelagem de processos da indústria de alimentos**. Orientador: Luis Antônio Minim. Coorientadores: Leonardo Bonato Felix, Renata Cássia Campos e Valéria Paula Rodrigues Minim.

A torra e o forneamento são responsáveis pelas transformações observadas em alguns alimentos, tais como café e pães. Dentre essas modificações, a cor é aquela de maior destaque pois consiste num indicador da evolução do processo e tem sido utilizada no desenvolvimento de ferramentas de controle. Dessa forma, a modelagem do escurecimento não enzimático decorrente dos processos de torra e forneamento possibilita o desenvolvimento de sistemas de visão computacional para acompanhamento e classificação desses processos. A modelagem fenomenológica, baseada na abordagem da cinética de reações, permitiu uma maior compreensão dessas mudanças de cor. Porém, os modelos resultantes apresentam grandes limitações de ordem prática para sua aplicação nos processos industriais, em especial no contexto da indústria de quarta geração (Indústria 4.0), a qual preconiza o uso de sistemas e equipamentos inteligentes. Nesse sentido, a adoção de técnicas de inteligência artificial (AI), em especial as redes neurais convolucionais (CNN), parece ser o caminho a ser seguido. Assim, neste trabalho foram introduzidas técnicas de modelagem por AI do escurecimento não enzimático decorrente dos processos de torra e forneamento de café e pães, respectivamente. A adoção de CNN com reduzido número de camadas convolucionais resultou numa redução no consumo de memória de mais de 90%. Além disso, o sistema híbrido formado por uma CNN e uma máquina de vetores de suporte (SVM) resultou na redução de 93% no tempo de convergência. Foram identificados mais 20% dos núcleos convolucionais seletivos a cores, o que evidencia a capacidade das CNN de extrair características de cores e utilizá-las para classificação das amostras. Ao classificar amostras de pães e café, as CNN apresentaram exatidão superior a 98,0% e 95%, respectivamente, superando arquiteturas tradicionais. A CNN também foi capaz de estimar o tempo necessário para o final do processo de torra de café com *Root Mean Square Error* (RMSE) de 0,4 min. As CNN se mostraram uma poderosa

ferramenta não invasiva e não destrutiva para modelagem do escurecimento não enzimático decorrente dos processos de torra e forneamento.

Palavras-chave: Torra. Forneamento. Redes Neurais Convolucionais. Escurecimento Não Enzimático. Aprendizado Profundo.

ABSTRACT

COTRIM, Weskley da Silva, D.Sc., Universidade Federal de Viçosa, August 2021. **Artificial intelligence applied to process modeling in the food industry**. Adviser: Luis Antônio Minim. Co-Advisers: Leonardo Bonato Felix, Renata Cássia Campos and Valéria Paula Rodrigues Minim.

Roasting and baking are responsible for the transformations observed in some foods, such as coffee and bread. Among these changes, color is the most prominent because it is an indicator of the evolution of the process and has been used in the development of control tools. Thus, the modeling of non-enzymatic browning resulting from roasting and baking processes enables the development of computer vision systems to track and classify these processes. Phenomenological modeling, based on the reaction kinetics approach, has allowed a better understanding of these color changes. However, the resulting models have major practical limitations for their application in industrial processes, especially in the context of the fourth generation industry (Industry 4.0), which advocates the use of intelligent systems and equipment. In this sense, the adoption of artificial intelligence (AI) techniques, especially convolutional neural networks (CNN), seems to be the way to be followed. Thus, in this work we introduced AI modeling techniques of non-enzymatic browning resulting from coffee and bread roasting and baking processes, respectively. The adoption of CNN with reduced number of convolutional layers resulted in a reduction in memory consumption of more than 90%. In addition, the hybrid system consisting of a CNN and a support vector machine (SVM) resulted in a 93% reduction in convergence time. More than 20% of the color-selective convolutional kernels were identified, which highlights the ability of CNNs to extract color features and use them for sample classification. When classifying bread and coffee samples, the CNNs showed accuracy of over 98.0% and 95%, respectively, outperforming traditional architectures. CNN was also able to estimate the time required for the end of the coffee roasting process with Root Mean Square Error (RMSE) of 0.4 min. CNNs proved to be a powerful non-invasive and non-destructive tool for modeling non-enzymatic browning arising from the roasting and baking processes.

SUMÁRIO

1. INTRODUÇÃO GERAL

Para muitas matérias-primas alimentícias, tais como cafés, castanhas, biscoitos e pães, a torra e o forneamento representam a última etapa de processamento para obtenção do produto acabado. Durante essa etapa, o calor induz uma série de transformações físico-químicas irreversíveis, as quais são responsáveis pelas características sensoriais do alimento pronto, tais como aroma, sabor, textura e cor. Dentre essas, as mudanças de cor são particularmente importantes, pois constituem um importante indicador dos estágios dos processos de torra e forneamento. Por esse motivo, em geral, o controle de tais processos é realizado mediante inspeção visual por operador treinado, o que pode levar a falhas no processo, com risco para a qualidade do produto final.

Para contornar tais limitações, tem-se empregado técnicas de modelagem do processo de escurecimento não enzimático decorrente da torra ou forneamento, com vistas a construção de sistemas de visão computacional (CVS) para controle automático do processo. Os primeiros modelos propostos se basearam no fato de que a formação da cor escura observada nos grãos de café ou castanhas, durante a torra, ou na crosta de pães e biscoitos, durante o forneamento, se deve aos compostos coloridos formados durante as reações de caramelização e de Maillard. Assim, numa abordagem fenomenológica baseada na cinética das reações, diversos modelos foram apresentados, utilizando leituras instrumentais da cor no espaço de cor CIELab como dados de entrada. Dentre os modelos apresentados, destacam-se pelo seu pioneirismo e qualidade dor resultados obtidos, os modelos cinéticos de ordem zero e primeira ordem. Todos esses modelos compartilhavam entre si a dependência da temperatura da superfície do alimento para estimativa da constante da taxa de reação. Em alguns casos também se observa a dependência da umidade da amostra, além da temperatura. Embora tais modelos tenham ampliado o entendimento da formação da cor durante a torra e forneamento das matérias-primas alimentícias, sua dependência de propriedades de difícil mensuração ou necessidade de medidas destrutivas limita seu uso prático no contexto industrial.

Posteriormente, a abordagem da cinética de reação foi substituída por modelagem por inteligência artificial com recursos tradicionais de aprendizado de máquina. Nesse contexto, técnicas de modelagem por redes neurais artificiais (ANN)

e máquinas de vetores de suporte (SVM) se destacaram. As ANN são treinadas, com base num banco de dados, para reconhecer padrões, imitando a capacidade humana de aprendizado, permitindo a tomada de decisões baseadas em "experiências" anteriores. Um dos principais atributos das ANN é a capacidade de lidar com as não linearidades de processos complexos como aqueles observados na indústria de alimentos, onde fatores como composição, fenômenos de transferência de calor e massa e reações químicas estão presentes, dificultando a construção de modelos de forma convencional. Já as SVM apresentam como principal característica sua grande capacidade discriminativa entre elementos de duas classes.

Na modelagem por aprendizado de máquina ainda foram utilizados os mesmos tipos de dados de entrada observados na abordagem cinética de reações, o que leva aqueles modelos a compartilharem das mesmas limitações impostas a estes. Como consequência, tais modelos nem sempre atendem os pressupostos da Quarta Revolução Industrial (Indústria 4.0), os quais preconizam o uso de equipamentos e sistemas inteligentes, nos quais a informação deve ser facilmente obtida e compartilhada de forma autônoma permitindo a rápida tomada de decisão, com minimização de perdas e custos.

Para superar tais limitações, uma nova abordagem para modelar o escurecimento não enzimático decorrente dos processos de torra e forneamento baseada em aprendizado profundo foi introduzida. Nesse cenário, a principal técnica adotada são as redes neurais convolucionais (CNN). As CNN possuem como principal característica a sua capacidade de extrair o mapa de características diretamente de dados complexos, como é o caso de imagens, sem a necessidade de intervenção humana. Tal habilidade é particularmente importante, pois atende os pressupostos da Indústria 4.0, possibilitando a construção de CVS inteligentes e autônomos, capazes de se integrarem ao ecossistema de uma fábrica inteligente. Além disso, as CNN compartilham com as ANN a sua capacidade de lidar com as não linearidades típicas dos processos da indústria de alimentos, o que as torna a ferramenta ideal para a modelagem do escurecimento não enzimático durante os processos de torra e forneamento.

Nesse contexto, na seção 2 será apresentada uma revisão sistemática da modelagem do processo de escurecimento não enzimático decorrente dos

processos de torra e forneamento utilizando a abordagem clássica por cinética de reações e a moderna abordagem por inteligência artificial. Na seção 3 será apresentado um modelo de rede neural convolucional com reduzido número de camadas convolucionais (Short-CNN). Aqui será demonstrada a capacidade das CNN de extrair características de cor diretamente das imagens e empregá-las na classificação dos diferentes estágios de forneamento de pães. Em seguida, na seção 4, será introduzido um modelo híbrido, o qual combina a capacidade das CNN de extrair características diretamente das imagens da crosta de pães, com a alta capacidade discriminativa das SVM, para classificação dos estágios de forneamento de pães. Já na seção 5 uma CNN com reduzido número de camadas convolucionais será aplicada na classificação de diferentes estágios de torra de café. Adicionalmente, será mostrada sua capacidade em estimar o tempo necessário para conclusão do processo de torra de uma amostra, baseado unicamente em imagens do processo. Finalmente, na seção 6 serão apresentadas as conclusões finais e futuros trabalhos.

2. ARTIGO 1 – USO DE INTELIGÊNCIA ARTIFICIAL NA MODELAGEM DO ESCURECIMENTO NÃO ENZIMÁTICO DE ALIMENTOS DECORRENTE DOS PROCESSOS DE TORRA E FORNEAMENTO

Neste artigo será apresentada uma revisão sistemática do processo de escurecimento não enzimático resultante da torra e forneamento de cafés, pães e biscoitos. Também será discutida a modelagem do processo na abordagem tradicional, baseada na cinética das reações de Maillard e de caramelização, e na moderna abordagem por inteligência artificial. Serão discutidas as vantagens da modelagem do escurecimento não enzimático durante a torra ou forneamento por inteligência artificial (AI), utilizando técnicas de aprendizado de máquina (*machine learning*) e de aprendizado profundo (*deep learning*). Também serão revisadas as principais técnicas de *machine learning* utilizadas para modelagem desse processo. No contexto da modelagem por técnicas de *deep learning* serão introduzidas as redes neurais convolucionais (CNN), com seus fundamentos teóricos e aplicações práticas. Também serão apresentados os principais modelos de AI aplicados a modelagem do processo do escurecimento resultante da torra ou forneamento. Finalmente, serão discutidos os desafios e oportunidades na aplicação da AI na modelagem das mudanças de cor em produtos alimentícios durante a torra ou forneamento.

Uso de Inteligência artificial na modelagem do escurecimento não enzimático de alimentos decorrente dos processos de torra e forneamento

Weskley da Silva Cotrim[1], Valéria Paula Rodrigues Minim[2], Leonardo Bonato Felix[3], Luis Antônio Minim[4,*]

[1] Universidade Federal de Mato Grosso, *Campus* Universitário do Araguaia, Instituto de Ciências Exatas e da Terra, Avenida Valdon Varjão, nº 6390. Barra do Garças - Mato Grosso, Brasil, CEP: 78600-000, weskleycotrim@ufmt.br, https://orcid.org/0000-0002-8190-9632

[2] Universidade Federal de Viçosa, Departamento de Tecnologia de Alimentos, *Campus* Universitário – Viçosa – Minas Gerais, Brasil, CEP: 36570-000, vprm@ufv.br, https://orcid.org/0000-0001-7143-2060

[3] Universidade Federal de Viçosa, Departamento de Engenharia Elétrica, *Campus* Universitário – Viçosa – Minas Gerais, Brasil, CEP: 36570-000, leobonato@ufv.br, https://orcid.org/0000-0002-6184-2354

[4] Universidade Federal de Viçosa, Departamento de Tecnologia de Alimentos, *Campus* Universitário – Viçosa – Minas Gerais, Brasil, CEP: 36570-000, lminim@ufv.br, https://orcid.org/0000-0002-1584-9117

*Corresponding author
Prof. Luis Antônio Minim
Universidade Federal de Viçosa, Department of Food Technology, *Campus* Universitário – Viçosa – Minas Gerais, Brasil, CEP: 36570-000, lminim@ufv.br

RESUMO

Os processos de torra e forneamento são cruciais para a qualidade final de produtos como pães e cafés, por induzirem importantes alterações no aroma, sabor, textura e cor. A mudança de cor destaca-se pelo seu papel com indicador do processo e por essa razão, a modelagem do escurecimento não enzimático decorrente da torra ou forneamento tem atraído a atenção de muitos pesquisadores. Tradicionalmente a modelagem do escurecimento não enzimático é realizada seguindo uma abordagem cinética. Entretanto, novas abordagens baseadas em técnicas de inteligência artificial, tais como aprendizado de máquina e aprendizado profundo, com destaque para as redes neurais convolucionais (CNN), têm sido introduzidas. A modelagem por inteligência artificial representa um novo paradigma e abre muitas possibilidades, especialmente no contexto da Indústria 4.0, superando abordagens tradicionais. Por outro lado, desafios como qualidade dos bancos de dados, tipos de modelos, aprendizagem incremental e tecnologias emergentes como o gêmeo digital representam oportunidades de pesquisa e desenvolvimento.

Palavras-chave: Convolutional neural network, deep learning, artificial intelligence.

1. Introdução

Torra e forneamento correspondem essencialmente à mesma operação unitária, sendo diferenciados apenas no uso popular. Em geral, o termo torra é aplicado ao processamento térmico de nozes, castanhas e outros grãos, como é o caso do café e do cacau (Bustos-Vanegas et al., 2018; Cha & Lee, 2020; Toker et al., 2020; Wani et al., 2017). Por outro lado, o termo forneamento é comumente aplicado aos produtos de panificação, tais como pães e biscoitos (Arepally et al., 2020; Purlis & Salvadori, 2009). Trata-se de um processo combinado de transferência de calor e de massa, no qual se utiliza ar quente e seco como veículo principal para transferência de calor e remoção de massa na forma de vapor d'água e outros gases (Bottazzi et al., 2012; Hernández et al., 2008). Embora a transferência de calor convectiva seja aquela de maior contribuição em processos realizados em fornos convencionais, também cabe mencionar que uma parcela significativa de calor é transferida por condução direta das paredes do forno e bandejas ou ainda por radiação térmica (Sruthi et al., 2021). Durante a torra ou forneamento, os alimento são aquecidos a temperaturas que variam de 150 a 300 °C, por um tempo que varia desde alguns segundos até cerca de uma hora, dependendo do produto e do objetivo a ser alcançado (Arepally et al., 2020; Sruthi et al., 2021; Zanoni et al., 1995).

Durante o processo, o ar seco aquecido entra em contato direto com a superfície do alimento transferindo calor para o mesmo e removendo a umidade superficial (Sruthi et al., 2021). Como consequência, observa-se grande evaporação de água, com formação de uma camada superficial com baixíssima umidade (Pour-Damanab et al., 2014). Internamente forma-se um gradiente de temperatura e umidade, o qual é responsável pela migração da água do interior para superfície e o derretimento das gorduras eventualmente presentes. Nessa etapa observa-se um aumento da plasticidade do material, o que se reflete na redução dos módulos de cisalhamento elástico e viscoso (G' e G'') (Laguna et al., 2013). Ao se atingir temperatura suficientemente alta, inicia-se o processo de gelatinização do amido, que é mais significativa em produtos de panificação, e a desnaturação e polimerização das proteínas (Sruthi et al., 2021). Com a continuidade no aumento da temperatura interna, inicia-se o processo de vaporização da água e perda de outros

gases, especialmente dióxido de carbono, oriundo das reações de caramelização e de Maillard (Arepally et al., 2020; Helou et al., 2016; X. Wang & Lim, 2014a).

Como consequência das reações de caramelização e da Maillard inicia-se a formação de uma série de compostos químicos, os quais são responsáveis pelas características de aroma, sabor e cor do produto acabado (Çelik & Gökmen, 2020; Getachew & Chun, 2018; Sacchetti et al., 2009). Tais componentes são particularmente importantes, pois são um indicativo de que o produto atingiu o estágio final de forneamento ou torra e impactam diretamente na decisão de compra do consumidor (Castro et al., 2017; Purlis, 2012). Dentre esses indicadores, destaca-se a cor, uma vez que a mesma é facilmente mensurável e correlaciona-se bem com outras propriedades do produto pronto (Baggenstoss et al., 2008; Chhanwal et al., 2011; Romani et al., 2012). Por essa razão, a cor é utilizada como indicador da evolução do processo de torra ou forneamento, seja por inspeção visual, seja no desenvolvimento de sistemas de controle de processo baseados em visão computacional (Demir et al., 2002; Sacchetti et al., 2009; X. Wang & Lim, 2014a).

Nesse sentido, diferentes abordagens já foram adotadas ao longo dos anos para modelagem do processo de escurecimento, com vistas ao desenvolvimento de sistemas de controle por visão computacional. As primeiras tentativas foram realizadas utilizando modelagem fenomenológica, com base no conhecimento da cinética das reações de Maillard e de caramelização (Purlis & Salvadori, 2009; Sacchetti et al., 2009; X. Wang & Lim, 2014a; Zanoni et al., 1995). Entretanto, tais modelos apresentam baixa capacidade de generalização, forte dependência de propriedades de difícil mensuração ou uso de sensores de alto custo de aquisição e manutenção, o que limita o seu uso prático pela indústria. Somam-se a isso os requisitos da quarta revolução (Indústria 4.0) pela qual passa a indústria de alimentos, a qual preconiza o uso de sistemas e equipamentos inteligentes (Li & Si, 2017; J. Wang et al., 2018; Zhong et al., 2017). Dessa forma, a adoção da inteligência artificial, em especial técnicas de aprendizado profundo, como é o caso das Redes Neurais Convolucionais (CNN), tem produzido bons resultados (Cotrim et al., 2020, 2021). Isso se deve ao fato de que as CNN são capazes de extrair, de forma autônoma, as características que serão utilizadas para alimentar o modelo a partir da informação bruta (Gu et al., 2018) e por lidarem bem com as não

linearidades presentes nos processos da indústria de alimentos (Putranto et al., 2015).

Até o momento, não existe uma revisão abrangente que trate da modelagem das mudanças de cor decorrente dos processos de torra e forneamento dos alimentos, o que exige um tentativa sistemática e exaustiva de resumir diferentes descobertas e tirar conclusões importantes. Nesse sentido é apresentada a seguir uma revisão sobre o processo de escurecimento não enzimático durante a torra e o forneamento de alimentos e a aplicação de técnicas de inteligência artificial na modelagem das mudanças de cor. Na seção 2 será apresentado um breve resumo do processo de torra de café e forneamento de pães e biscoitos. No tópico seguinte é abordado o histórico da modelagem tradicional desses processos. Em seguida, na seção 3, será discutida a modelagem desses processos baseados na abordagem tradicional e por técnicas de inteligência artificial. Serão descritas as duas principais técnicas de aprendizado de máquina com ampla aplicação na modelagem de Sistemas de Visão Computacional (CVS) para acompanhamento do escurecimento decorrente do processamento térmico: as redes neurais artificiais (ANN) e as máquinas de vetores de suporte (SVM). Na seção 4 será introduzida a principal técnica de aprendizagem profunda utilizada na modelagem de CVS: as redes neurais convolucionais (CNN). Neste tópico são detalhados os fundamentos teóricos das CNN. Na seção 5 serão apresentadas as principais aplicações da inteligência artificial no campo do aprendizado de máquina (ML) e do aprendizado profundo (DL) para modelagem do processo de escurecimento em produtos forneados e torrados. Finalmente, na seção 6, serão discutidos os desafios e oportunidades na aplicação das técnicas de aprendizado profundo, em especial as CNN, na modelagem do processamento térmico de alimentos.

2. Processos de torra e forneamento

2.1. Processo de torra de café

A torra do café ocorre em temperaturas próximas aos 200 °C, podendo ser realizada em modo contínuo ou em batelada. Quanto aos equipamentos, podem ser do tipo leito fluidizado, no qual a transferência de calor ocorre predominantemente por via convectiva, ou torradores rotativos, onde a transferência de calor condutiva predomina (Vargas-Elías et al., 2016). É durante a torra que os compostos

responsáveis pelo aroma, corpo, cor e acidez da bebida final são formados, o que torna essa etapa crucial para a qualidade final percebida pelo consumidor (Baggenstoss et al., 2008; Romani et al., 2012; X. Wang & Lim, 2014a).

A torra do café ocorre em três etapas básicas: *i)* Desidratação *ii)* torra propriamente dita e *iii)* resfriamento. A primeira etapa da torra, de caráter eminentemente endotérmico, ocorre quando os grãos atingem temperatura entre 90 °C e 120 °C. Nessa condição o grão começa a perder massa na forma de vapor d'água. É nessa etapa que se observam as primeiras modificações na cor dos grãos, os quais passam de verde claro para amarelo ou alaranjado. Quando a temperatura interna do grão atinge valores na faixa de 120 °C a 150 °C ocorre a liberação dos primeiros componentes de aroma, os quais são responsáveis pelo aroma característico de pão cozido. Essa etapa é caracterizada pela expansão dos grãos causada pelo aumento da pressão interna devido a vaporização da água no interior do grão e início da liberação do dióxido de carbono, fruto das alterações químicas que se iniciaram (X. Wang & Lim, 2014b).

A torra propriamente dita acontece quando o grão atinge temperatura interna de 180 °C e é marcada pela ocorrência do fenômeno conhecido como o "primeiro *crack*". Tal fenômeno é caracterizado pelo rompimento dos grãos devido ao aumento da pressão interna causado pelo acúmulo de vapor d'água e gás carbônico (X. Wang & Lim, 2017). É também nessa etapa que se inicia a reação de caramelização dos açúcares presentes no grão, a qual é facilmente identificada pela cor marrom observada. Ao final do "primeiro *crack*", quando a temperatura interna do grão se situa na faixa de 205 °C a 220 °C, cerca de 50% dos açúcares passaram pelo processo de caramelização. Nesse momento, se observa a cor marrom moderada característica e bom balanço entre acidez e teor de açúcares. Também se observa grande liberação de componentes responsáveis pelo aroma e sabor característico do café. Ao atingir cerca de 230 °C tem-se o que é conhecido como "segundo *crack*", fenômeno decorrente do rompimento estrutural do grão em função da ação do calor sobre as estruturas organizadas de celulose. Também é observado aumento do brilho superficial do grão devido a liberação de óleo. Caso o processo de torra prossiga, ocorrerá aumento da temperatura interna do grão para cerca de 240 °C. Nessa condição começa haver rápida degradação dos componentes de aroma e formação do aroma queimado devido a carbonização das estruturas celulósicas. Em

caso de torra mais intensa (240 °C a 265 °C) a carbonização se intensifica com o predomínio de tons amargos de queimado (Virgen-Navarro et al., 2016).

Ao final da etapa de torra propriamente dita, inicia-se a etapa de resfriamento a qual visa interromper as alterações físico-químicas da etapa anterior e é geralmente realizada pela passagem de um fluido a baixas temperaturas (ar ou água) pelos grãos (Czech et al., 2016).

2.2. Processo de forneamento de pães e biscoitos

A produção de pães e biscoitos envolve uma etapa de cozimento da massa, denominada forneamento, para obtenção das características sensoriais desejadas. Durante o forneamento, o produto passa por uma série de transformações na sua estrutura e aparência em função da perda de massa, na forma de água evaporada, e escurecimento na superfície decorrente das reações de Maillard e de caramelização (L. Zhang et al., 2016). As modificações na cor são especialmente importantes, pois são um indicador da qualidade do processo e elemento fundamental de aceitação do produto pelo consumidor (Purlis, 2012).

As reações de Maillard e de caramelização são as principais responsáveis pela formação da cor marrom característica em pães e biscoitos (Arepally et al., 2020; Purlis & Salvadori, 2009). Tal reação inicia-se quando a temperatura da superfície atinge valores na faixa de 105 °C a 115 °C, o que ocorre quando se forma uma crosta com baixa umidade. Nessas condições observa-se o consumo de carboidratos pela reação, com consequente produção de substâncias coloridas (melanoidinas). É também nessa etapa que se observa a liberação dos compostos responsáveis pelo aroma (Vanin et al., 2009).

3. Abordagem cinética na modelagem das mudanças de cor no processamento térmico

Em produtos alimentícios, tais como cafés, pães e biscoitos, a cor é elemento essencial no processo de avaliação da qualidade por estar associada com a formação de compostos que afetam o aroma e sabor do produto final (Bertrand et al., 2018). Sua formação está ligada a presença de carboidratos e aminoácidos livres, os quais quando submetidos a aquecimento em baixa umidade, participam das reações de caramelização e de Maillard (Kocadağli & Gökmen, 2016).

Assim, os primeiros modelos propostos para predição do escurecimento não enzimático em produtos submetidos a aquecimento tiveram como base, evidências de que as reações de Maillard e de caramelização seguem cinética de ordem zero ($n = 0$) ou de ordem "mista" ($0 < n < 1$) (Buera et al., 1987; Stamp & Labuza, 1983).

Entretanto, pesquisas posteriores indicam que essas reações apresentam comportamento mais complexo, onde modelos multirresposta parecem ser mais adequados (Kocadağli & Gökmen, 2016; Quintas et al., 2007). Por esse motivo, a modelagem do escurecimento decorrente do processamento térmico de alimentos ainda persiste como um grande desafio a ser vencido. Como será visto, algumas tentativas foram feitas para a construção de sistemas de controle baseados nas mudanças de cor, utilizando a abordagem da cinética das reações. Porém, tais modelos sofrem com a baixa capacidade de generalização ou dependência de variáveis de difícil mensuração, tais como temperatura ou umidade do produto, o que tem limitado a sua aplicação prática.

3.1. Modelagem da mudança de cor durante o processo de torra de café

A cor é um dos principais indicadores do grau de torra do café. Vários estudos demonstraram existir correlação entre as mudanças de cor e as alterações de umidade, volume, massa, teor de compostos voláteis, teor de dióxido de carbono no grão e o tempo de processo de torra (Baggenstoss et al., 2008; Bicho et al., 2012; Kim et al., 2018; Romani et al., 2012; Vargas-Elías et al., 2016; Virgen-Navarro et al., 2016; X. Wang & Lim, 2017).

Nesse contexto, alguns trabalhos foram desenvolvidos com o objetivo de se obter um modelo preditivo do grau de torra em função da cor do grão, sempre utilizando como referência o espaço de cor CIELab (Sacchetti et al., 2009; X. Wang & Lim, 2014a). Nessa escala de cor, o índice de luminosidade (L*) decresce continuamente com o aumento no tempo de torra, enquanto os índices de vermelho e verde (a*) e amarelo e azul (b*) aumentam nos instantes iniciais da torra, voltando a decrescer e se estabilizar em um dado valor com o prosseguimento do processo (Baggenstoss et al., 2008; X. Wang & Lim, 2014a). Por esse motivo, o índice de luminosidade tem sido utilizado na geração de modelos matemáticos empíricos ou fenomenológicos (Bicho et al., 2012).

Os principais modelos encontrados na literatura se baseiam no pressuposto de que a variação do índice de luminosidade segue comportamento de reação elementar. Dessa forma, vemos o processo de escurecimento dos grãos de café durante a torra ser representado por modelos de segunda ordem (Sacchetti et al., 2009) ou por modelos de primeira ordem (X. Wang & Lim, 2014a). Tais modelos permitiram um melhor entendimento do processo de escurecimento dos grãos de café durante a torra. Entretanto, esses mesmos modelos apresentam desvios significativos nas etapas finais do processo ou em temperatura mais elevadas, o que é crítico para a sua confiabilidade e aplicação prática.

Embora modelos cinéticos de primeira e segunda ordem predominem na modelagem do escurecimento de grãos de café durante a torra, recentemente foi demonstrado que a ordem do modelo é fortemente influenciada pela temperatura do ar no interior do equipamento (Chindapan et al., 2019). Nesse caso verificou-se uma mudança de um modelo de primeira ordem para um de ordem 3 ao se elevar a temperatura do ar para 250 °C, o que implica numa maior sensibilidade do modelo para mudanças nos valores de luminosidade (L*).

3.2. Modelagem da mudança de cor durante o processo de forneamento de pães e biscoitos

Semelhante ao que ocorre com o café durante a torra, no forneamento de pães e biscoitos se observa uma intensa modificação da cor ao longo do tempo de exposição ao calor. Entretanto, ao contrário do que ocorre com o café, onde a mudança de cor ocorre na superfície e no interior do grão, no processo de forneamento a mudança de cor é predominantemente superficial e está associada com a formação de uma crosta de baixa umidade (Mundt & Wedzicha, 2007; Purlis & Salvadori, 2009). Além disso, em estudo de forneamento de biscoitos foi demonstrado que a formação da cor é dependente da temperatura superficial da amostra (Shibukawa et al., 1989). Tal achado foi importante, pois permitiu que se adotasse o modelo de Arrhenius como base para a estimativa das constantes de reação.

Com base nesse pressuposto, os primeiros trabalhos apresentaram modelos baseados na cinética de reação de primeira ordem para predição da cor da crosta do pão em função do tempo, utilizando os três índices de cor (L*, a* e b*) na escala

CIELab (Zanoni et al., 1995). Além da cor, o modelo também dependia da temperatura superficial da amostra para estimativa da constante de reação, o que impunha uma limitação ao seu uso prático. Na mesma direção, foi adotado modelo de primeira ordem para estimativa das mudanças de cor durante o forneamento de biscoitos com base na luminosidade das amostras (L*) (Broyart et al., 1998). Aqui, a constante de reação foi estimada com base também no teor de umidade da amostra, além da temperatura.

Mais tarde, verificou-se que a temperatura e o tempo de processo são os dois principais fatores que influenciam a formação de cor da crosta de pães durante o forneamento (Ramírez-Jiménez et al., 2000). Também ficou claro, assim como se verificou durante o forneamento de biscoitos, que a cor da crosta de pães é altamente dependente da temperatura superficial da amostra (Wählby & Skjöldebrand, 2002). Além disso, diferentes abordagens para a modelagem do escurecimento durante o forneamento de pães têm sido propostas. Numa delas, foi proposto um modelo de escurecimento da crosta de pães e biscoitos baseado na relação exponencial entre a cor e o conteúdo de melanoidinas resultante da reação de Maillard (Hadiyanto et al., 2007). Noutra, o escurecimento da crosta dos pães foi modelado em função da perda de massa durante o forneamento (Purlis & Salvadori, 2007).

Entretanto, tais abordagens foram logo abandonadas, e a abordagem cinética foi retomada com ajustes. Nesse contexto foi proposto um modelo de primeira ordem baseado nas mudanças de cor da crosta de pães, expresso como luminosidade, porém obtido da conversão direta da cor obtida em imagens no espaço de cor RGB (Purlis & Salvadori, 2009). Nesse caso, a constante de reação foi estimada com base na temperatura e na atividade de água da superfície da amostra, o que acrescenta uma camada extra de complexidade ao modelo. Como resultado, o modelo apresentou boa capacidade preditiva. Entretanto, a dependência de variáveis de difícil mensuração torna inviável seu uso fora do ambiente acadêmico.

Mais recentemente, combinou-se o modelo proposto por Zanoni et al. (1995) com a *"Spatial Reaction Engineering Approach"* (S-REA), adotando a premissa de Purlis e Salvadori (2009) de que o conteúdo de água na crosta também influencia a velocidade de escurecimento da amostra (Putranto et al., 2015). A S-REA combina os princípios de engenharia das reações químicas com as equações governantes de

transferência de calor e massa para modelagem da cinética de desidratação superficial em alimentos (Putranto et al., 2011, 2015). Em tal abordagem, adotou-se cinética de primeira ordem na sua forma integral para descrever o escurecimento da superfície da amostra. Embora mais complexo, o modelo não apresentou vantagens significativas, uma vez que ainda foram observadas regiões com grande desvio entre o valor predito e o valor medido experimentalmente.

4. Modelagem do escurecimento não enzimático por inteligência artificial

A modelagem do escurecimento não enzimático representa um grande desafio para a indústria de alimentos devido a complexidade das reações químicas envolvidas. Associado a isso, a heterogeneidade de distribuição espacial dos componentes, bem como a complexidade da matéria-prima constituem num entrave extra para a modelagem fenomenológica de tal processo (Sruthi et al., 2021). Como resultado, a construção de modelos mecanísticos representativos da variação temporal da cor torna-se demasiadamente complexa em razão das não linearidades observadas (Demir et al., 2002; Purlis, 2011; L. Zhang et al., 2016).

Outra limitação dos modelos fenomenológicos reside na aplicabilidade prática dos mesmos. Como visto, os modelos preditivos do escurecimento não enzimático descritos na literatura apresentaram razoável capacidade de ajuste aos dados experimentais, contribuindo para o entendimento do fenômeno. Entretanto, os mesmos apresentam baixa utilidade prática, uma vez que dependem de informações pouco práticas no dia a dia de uma indústria, tais como a inclusão de sensores de temperatura ou umidade na superfície de amostras de pães ou no interior de gãos de café (Purlis & Salvadori, 2009; Putranto et al., 2015; X. Wang & Lim, 2014a).

Soma-se a isto o fato de que tais modelos não atendem os pressupostos da Quarta Revolução Industrial, também conhecida como Indústria 4.0, os quais preconizam o uso de equipamentos e sistemas inteligentes. Em tais equipamentos e sistemas, a informação deve ser facilmente obtida e compartilhada de forma autônoma permitindo a rápida tomada de decisão, com minimização de perdas e custos (Li & Si, 2017; J. Wang et al., 2018; Zhong et al., 2017). Fábricas inteligentes representam um novo paradigma de produção onde os equipamentos estão totalmente conectados entre si, monitorados por sensores e controlados por inteligência computacional avançada visando aumentar a qualidade dos produtos, a

produtividade do sistema e a sustentabilidade, enquanto reduz custos (J. Wang et al., 2018). Nesse contexto, o uso de algoritmos de inteligência artificial pode ser uma alternativa útil para a construção de modelos de controle de processo, uma vez que as áreas de aprendizado de máquina (*machine learning*) e aprendizado profundo (deep learning) são consideradas pilares para o desenvolvimento da Indústria 4.0 (Bortolini et al., 2017; Kotsiopoulos et al., 2021).

Por definição, Inteligência artificial (IA) é a habilidade de um computador, ou outra máquina, realizar atividades que requerem Inteligência humana. O uso de algoritmos de inteligência artificial, ou inteligência computacional, tornou-se popular na modelagem de processos industriais devido a sua robustez, simplicidade na formulação, facilidade de modelagem e capacidade de adaptação (Lee et al., 2018). Além disso, a IA não requer conhecimento prévio do comportamento do processo, o qual muitas vezes é difícil de ser obtido, geralmente apresentando comportamento não linear (Bini, 2018). Os principais algoritmos de inteligência artificial utilizados para modelar processos industriais são as redes neurais artificiais (ANN) (Delgado et al., 2016), a lógica fuzzy (Chaibakhsh et al., 2011), o algoritmo genético (AG) (McCall, 2005), as máquinas de vetores de suporte (Battineni et al., 2019) e os sistemas híbridos (H. M. Wei et al., 2009; N. C. Wei et al., 2008).

4.1. Redes neurais artificiais

Dentre os algoritmos de aprendizagem de máquina, as redes neurais artificiais são os mais populares, devido a sua capacidade de aprendizado baseado em um conjunto de dados prévios do processo (Varol et al., 2018). Aliado a isso, as ANN apresentam grande precisão e consistência nas respostas do modelo, mesmo quando mudanças ocorrem no processo (J. M. Ali et al., 2015). Formalmente, redes neurais artificiais (ANN) são definidas como sistemas computacionais constituídos por um número de nós simples (neurônios) e altamente interconectados, os quais processam a informação em caráter dinâmico, paralelizado e distribuído, a partir de dados externos (Bai et al., 2018; Gonçalves et al., 2005; Willis et al., 1991). Estes elementos são conectados por meio de sinapses artificiais, simbolizadas por uma matriz de pesos os quais podem ser ajustados através de um processo de aprendizagem (treinamento) (Cao et al., 2018). ANN do tipo Multilayer Perceptron (MLP) são compostas por uma camada de dados de entrada (variáveis

independentes), uma ou mais camadas internas (ocultas) onde os dados são processados e uma camada de saída (variáveis dependentes). O objetivo de uma rede neural artificial é simular a capacidade humana de aprendizagem a partir de exemplos, permitindo a tomada de decisão pela máquina (Delgado et al., 2016). Para isso, cada neurônio artificial é responsável por receber uma série de dados de entrada (inputs), estabelecer pesos de ponderação (*weight factors*), nos quais a informação de aprendizagem é armazenada, processar a soma dos sinais de entrada (*transfer functions*) e produzir o resultado de saída (*outputs*) (J. M. Ali et al., 2015; Sudha et al., 2016).

Os tipos de redes neurais artificiais mais comumente utilizadas na modelagem de processos industriais são as *Feed Forward Neural Networks* (FFN), *Internally Recurrent Net* (IRN), *Externally Recurrent Net* (ERN), *Radial Basis Function Networks* (RBFN) e as *Shape-Tuneable Neural Networks* (MNN) (J. M. Ali et al., 2015). Diversos trabalhos têm reportado a aplicação de redes neurais artificiais em processos industriais, dentre eles destaca-se o uso na construção de modelos preditivos (León-Roque et al., 2016; Patan, 2018), otimização de processos (Sudha et al., 2016) e classificação de produtos (Debska & Guzowska-Świder, 2011).

Na indústria de alimentos, redes neurais têm sido aplicadas com sucesso na construção de modelos para predição da concentração de pigmentos em tratamento de efluentes (I. Ali et al., 2018), na modelagem de processo de esterilização de alimentos enlatados (Gonçalves et al., 2005), na classificação de carne suína fresca, congelada e descongelada e carne deteriorada e do tipo de músculo de origem (Górska-Horczyczak et al., 2017).

4.2. Máquinas de vetores de suporte

Ainda dentro da categoria de aprendizagem de máquina temos as Máquinas de Vetores de Suporte (SVM) (Cortes & Vapnik, 1995). As SVM são classificadores binários supervisionados, capazes de distinguir duas classes linearmente separáveis mediante uso de um hiperplano de decisão (Nalepa & Kawulok, 2019). Trata-se de um poderoso classificador, com aplicações nas áreas industriais (Tseng et al., 2016), de saúde (Battineni et al., 2019; Razzaghi et al., 2016) e segurança (Michele et al., 2019), dentre outras. Baseia-se no estabelecimento de dois vetores paralelos, equidistantes do hiperplano de decisão e posicionados sobre os elementos de cada

classe mais próximos do hiperplano de decisão, numa condição tal que a distância entre os dois vetores seja máxima (Cotrim et al., 2021).

As SVM foram originalmente projetadas para lidar com conjunto de dados totalmente linearmente separáveis. Entretanto, nem sempre existe uma fronteira clara entre dois conjuntos de dados no mundo real. Dessa forma, algumas modificações foram incorporadas ao modelo para permitir sua aplicação. A primeira modificação consiste na adoção da chamada margem suave (*soft margin*), a qual se baseia na aplicação de uma penalidade ao modelo, permitindo que o mesmo cometa alguns erros de classificação, porém aumente sua capacidade de generalização (Cortes & Vapnik, 1995). Isso é particularmente útil ao se lidar com a presença de *outliers* ou ruídos no conjunto de dados (Muller et al., 2001).

Outra modificação importante foi introduzida para lidar com dados que não são satisfatoriamente separáveis por um hiperplano e ficou conhecida como truque do núcleo (Nalepa & Kawulok, 2019). O truque do núcleo consiste em se aplicar uma função núcleo ao conjunto de dados original e convertê-lo para um espaço de maior dimensão, denominado espaço de características, no qual seja possível separar linearmente as classes (Battineni et al., 2019). As funções núcleo mais utilizadas são as Polinomiais, a Função Gaussiana de Base Radial (RBF) e a Sigmoidal (Tomar & Agarwal, 2015).

Por se tratar de um classificador binário, para aplicar as SVM a problemas multiclasse é necessário adotar uma das seguintes estratégias. Na primeira, do tipo um-versus-um (OVO), realiza-se a separação de pares de classes, ignorando os elementos pertencentes às demais classes. Nesse caso, para um problema de classificação com n classes, são necessários $n(n-1)/2$ classificadores. No segundo caso, o problema pode ser resolvido seguindo a abordagem um-versus-todos (OVA), no qual uma classe é separada de todas as outras. Assim, para um problema com n classes são necessários n hiperplanos de separação (Nashat & Abdullah, 2010).

5. Redes neurais convolucionais

Embora as redes neurais artificiais (ANN) e as máquinas de vetores de suporte (SVM) desempenhem um papel muito importante na modelagem de processos industriais, sua dependência da intervenção humana na seleção das

características que alimentam o modelo limitam seu uso no contexto da industria 4.0. Mais recentemente, dentro do conceito de aprendizagem profundo, foram introduzidas as redes neurais convolucionais (CNN), as quais permitem o reconhecimento de padrões complexos em grandes volumes de dados. Tal característica possibilita a modelagem de sistemas complexos sem a necessidade de intervenção humana na seleção das características que alimentam o modelo (He et al., 2016; Krizhevsky et al., 2012; Simonyan & Zisserman, 2015; Szegedy et al., 2015).

A primeira rede neural convolucional (CNN), conhecida como LeNet-5, foi proposta por LeCun e colaboradores para reconhecimento de caracteres escritos à mão (Le Cun et al., 1989). A LeNet-5 é considerada um marco para o desenvolvimento do aprendizado profundo (*deep learning*) (Gu et al., 2018; H. Wu & Zhao, 2018), por combinar a extração de padrões complexos presentes em imagens, sem a necessidade de intervenção humana, com o poder de classificação das redes MLP (Lecun et al., 1998; J. Wang et al., 2018).

Embora a arquitetura da LeNet-5 tenha sido considerada revolucionária à época, o seu alto custo computacional para implementação limitava sua utilização prática. O interesse pelas CNN só foi retomado em 2012, com o avanço da computação em GPU (*Graphics Processing Unit*) (Schmidhuber, 2015). Nesse contexto a nova arquitetura, denominada AlexNet, superou com larga margem os modelos da época no ImageNet Large Scale Visual Recognition Challenge (ILSVRC 2012 contest) (Krizhevsky et al., 2012). Desse momento em diante, várias arquiteturas foram propostas, sempre com foco no aumento da performance da CNN. Embora atualmente exista um número considerável de arquiteturas, merecem destaque pelo seu pioneirismo a VGGNet (Simonyan & Zisserman, 2015), a GoogLeNet (Szegedy et al., 2015) e a ResNet (He et al., 2016).

O poder das CNN reside no fato de ser uma inteligência orientada por dados a qual tem o poder de modelar relações complexas multivariadas entre bancos de dados sem a necessidade de conhecimento prévio da relação entre os mesmos (J. Wang et al., 2018). Tal premissa está completamente alinhada com o conceito de Indústria 4.0, onde processos complexos são modelados com base em grandes bancos de dados, de onde se extrai padrões ocultos ou imperceptíveis aos olhos humanos (Zhong et al., 2017).

5.1. Topologia básica de uma CNN

Embora atualmente existam várias arquiteturas diferentes de redes neurais convolucionais, alguns componentes são comuns a todas elas. Existem basicamente três tipos de camadas na rede neural convolucional. A camada de convolução propriamente dita, a camada de *pooling* e a camada totalmente conectada (*fully connected*). A camada convolucional é composta por uma série de núcleos ou filtros de convolução (*kernels*) os quais são utilizados para gerar as diferentes matrizes com mapas de características (*feature maps*) (Guo et al., 2016; Unnikrishnan et al., 2018). Cada neurônio de cada tensor do mapa de características está diretamente conectado a uma dada região do tensor da camada anterior, tal que essa região é definida pelo tamanho do núcleo e do passo (*stride*) de convolução (Gu et al., 2018). A Equação 1 é uma representação do processo de convolução, onde um tensor 4x4 é convoluída com um núcleo de convolução 2x2, produzindo um tensor 3x3.

$$\begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}_{4x4} \circledast \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}_{2x2} = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 1 & 2 \\ 1 & 3 & 1 \end{bmatrix}_{3x3} \qquad (Eq.\,1)$$

Observe que, para cada filtro convolucional, é gerado um novo mapa de características, os quais apresentam ordem menor que aquela da camada anterior (H. Wu & Zhao, 2018). Cada um dos elementos do tensor 4x4 representam os neurônios daquela camada. Assim como os elementos do tensor 3x3 também representam os neurônios do mapa de características gerado nessa nova camada. O elemento que une os dois tensores é o núcleo de convolução o qual é compartilhado por todos os elementos do tensor do mapa de características extraído (Gu et al., 2018). É este núcleo convolucional que deverá ser treinado pelo algoritmo de retropropagação. Matematicamente, o valor de cada característica z_{ijk}^l na posição (i, j), no *k*-ésimo tensor do mapa de características, na *l*-ésima camada convolucional, pode ser obtido conforme Equação 2.

$$z_{ijk}^l = \boldsymbol{W}_k^{l^T} \boldsymbol{X}_{ij}^l + b_k^l \qquad (Eq.\,2)$$

onde $\boldsymbol{W}_k^l$ é o tensor dos pesos (núcleo convolucional) de ponderação do *k*-ésimo filtro de convolução na *l*-ésima camada convolucional, b_k^l o desvio do *k*-ésimo núcleo

de convolução na *l*-ésima camada convolucional e $\boldsymbol{X}_{ij}^{l}$ o fragmento da matriz de entrada centrado na posição (i,j) (H. Wu & Zhao, 2018).

Semelhante ao que ocorre na rede neural artificial (ANN), aplica-se uma função de ativação a cada elemento do tensor $\boldsymbol{Z}_{ijk}^{l}$ para introduzir não linearidades e detectar características não lineares. Matematicamente, a função de ativação na rede neural convolucional pode ser dada por (Equação 3) (Cotrim et al., 2021):

$$a_{ijk}^{l} = a\big(z_{ijk}^{l}\big) \qquad (Eq.\,3)$$

A função de ativação aqui mencionada pode ser do tipo sigmoidal, tangente hiperbólica (Gu et al., 2018) ou, como é mais comum, do tipo ReLU (*Rectified Linear Unit*) (Nair & Hinton, 2010).

A camada de *"pooling"* visa obter invariância de deslocamento da característica extraída, mediante redução na resolução (dimensão) do tensor do mapa de características (Gu et al., 2018). A Equação 4 exemplifica o processo de *"pooling"* utilizando a função *"Max Pooling"*. Note que ao tensor 4x4 o *"pooling"* foi aplicado num fragmento (*patch*) 2x2 para a obtenção do valor correspondente no tensor resultante 2x2.

$$max\left(\begin{bmatrix} 2 & 2 & 2 & 3 \\ 2 & 1 & 1 & 1 \\ 1 & 3 & 1 & 4 \\ 4 & 2 & 1 & 2 \end{bmatrix}_{4x4} \Big/ _{2x2}\right) = \begin{bmatrix} 2 & 3 \\ 4 & 4 \end{bmatrix}_{2x2} \qquad (Eq.\,4)$$

Existem outras funções para realização do *"pooling"* tais como *"Average Pooling"*, *"L_p Pooling"*, *"Mixed Pooling"*, *"Sthochastic Pooling"*, dentre outras. De forma genérica, a operação de *"pooling"* pode ser representada matematicamente como (Equação 5):

$$y_{ijk}^{l} = pool\big(a_{mnk}^{l}\big), \forall\,(m,n) \in \mathcal{R}_{ij} \qquad (Eq.\,5)$$

sendo y_{ijk}^{l} o *k*-ésimo elemento do tensor da *l*-ésima camada de *"pooling"* próximo a posição (i,j). O empilhamento de sucessivas camadas de convolução e de *"pooling"* é responsável por extrair do banco de dados características de alto nível, permitindo que tais informações possam ser processadas pela rede neural na etapa *"fully-connected"* de maneira que se obtenha a solução desejada. Assim como ocorre na ANN, a última camada da CNN é responsável por produzir o tensor com os dados de

saída da rede os quais deverão ser comparados com um tensor de dados experimentais. Em problemas de classificação é comum que se utilize a função Softmax (Equação 6) como função de ativação na última camada da rede neural (H. Wu & Zhao, 2018).

$$\sigma\left(z_{ij}\right)_k = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}, \qquad j = 1, 2, \ldots, k \qquad (Eq.\,6)$$

O treinamento de uma rede neural convolucional é um problema de otimização global, portanto, é necessário que se estabeleça uma função custo a ser minimizada. Para uma relação de N pares de dados de entrada-saída desejada $\left\{\left(\boldsymbol{x}^{(n)}, \boldsymbol{d}^{(n)}\right); n \in [1, 2, \ldots, N]\right\}$, sendo $\boldsymbol{x}^{(n)}$ o *n*-ésimo dado de entrada e $\boldsymbol{d}^{(n)}$ o correspondente *n*-ésimo dado de saída desejado, com $\boldsymbol{y}^{(n)}$ representando o *n*-ésimo dado de saída da CNN, $\boldsymbol{\theta}$ um dado parâmetro a ser ajustado, a função custo a ser minimizada será a função perda ($\mathcal{L}$), a qual pode ser generalizada como (Equação 7) (Cotrim et al., 2021):

$$\mathcal{L} = \frac{1}{N} \sum_{n=1}^{N} \ell\left(\boldsymbol{\theta};\, \boldsymbol{d}^{(n)}, \boldsymbol{y}^{(n)}\right) \qquad (Eq.\,7)$$

5.2. Funções de ativação

Funções de ativação são essenciais numa rede neural, as quais são responsáveis pela introdução das não linearidades no modelo. Dentre as principais funções de ativação, destacam-se a função sigmoide, a tangente hiperbólica e a ReLU (Gu et al., 2018). A função sigmoide é responsável por converter os dados de entrada em um sinal de saída no intervalo (0, 1) (N. Liu et al., 2015) (Equação 8):

$$f(z_i) = \frac{1}{1 + e^{-z_i}} \qquad (Eq.\,8)$$

A função tangente hiperbólica (Equação 9), por outro lado, possui intervalo de saída (-1, 1) sendo representada matematicamente na forma (Debska & Guzowska-Świder, 2011; Funes et al., 2015):

$$f(z_i) = \frac{e^{z_i} - e^{-z_i}}{e^{z_i} + e^{-z_i}} \qquad (Eq.\,9)$$

Para as CNN a função de ativação não saturada mais comum é a *Rectified Linear Unit* (ReLU), a qual é definida como (Equação 10):

$$f(z_{ijk}) = max(z_{ijk}, 0) \qquad (Eq.\,10)$$

Como essa função basicamente converte todos os valores negativos para zero e mantêm os valores positivos, a mesma apresenta velocidade de processamento muito superior a aquela observada para as funções sigmoidal e tangente hiperbólica. Existem outras formas derivadas da função ReLU, tais como a função *Leakey* ReLU (LReLU), a função *"Exponential Linear Unit"* (ELU), dentre outras (Gu et al., 2018).

6. Aplicações de Inteligência artificial na modelagem do escurecimento não enzimático durante a torra e forneamento

Ao longo dos últimos anos tem crescido o interesse dos pesquisadores pelo uso da inteligência artificial na modelagem de processos da indústria de alimentos. Tal interesse também é observado na modelagem do escurecimento causado pelo processamento térmico durante a torra ou forneamento. Na Tabela 1 encontram-se sumarizados os principais trabalhos nos quais foram aplicadas técnicas de inteligência artificial na modelagem desses processos. As técnicas de aprendizado de máquina, tais como redes neurais artificiais (ANN), máquinas de vetores de suporte (SVM), general regression neural network (GRNN) e Adaptive network-based fuzzy inference system (ANFIS) foram as primeiras a serem adotadas (Broyart & Trystram, 2003; Leme et al., 2019; Nashat & Abdullah, 2010; Romani et al., 2012; Virgen-Navarro et al., 2016). Mais recentemente, tem crescido o interesse dos pesquisadores por técnicas de aprendizado profundo, em especial aplicações com as redes neurais convolucionais (CNN) (Chen et al., 2019; Cotrim et al., 2020, 2021; Hakim et al., 2020).

Tabela 1. Aplicações de *machine learning* e *deep learning* na construção de modelos para acompanhamento e predição das mudanças de cor durante os processos de torra e forneamento de alimentos.

Produto	Applicação	Modelo	Resultado	Referência
Biscoito	Predição da cor de forneamento	Dynamic model ANN	RMSE de 9,39	Broyart & Trystram, 2003
Biscoito	Classificação do estágio de forneamento	SVM and Wilk's λ analysis	Exatidão de 87,25%	Nashat & Abdullah, 2010
Café	Classificação do estágio de torra e predição do tempo de torra	ANN and GRNN	R^2 de 0,982 e RMSE de 1,786	Romani et al., 2012
Café	Classificação do estágio de torra	ANFIS	$R^2 > 0,98$ e RMSE $< 0,002$	Virgen-Navarro et al., 2016
Café	Predição de cor de torra	ANN	R^2 de 0,99	Leme et al., 2019
Pão	Geração de imagens RGB fotorrealísticas da crosta de pães de forma preditiva	GAN	Imagens geradas com sucesso	Chen et al., 2019
Pão	Classificação do estágio de forneamento (sete classes)	Short-CNN	Short-CNN apresentou núcleos seletivos às cores presentes na crosta do pão, exatidão de 98,8%,	Cotrim et al., 2020
Café	Classificação do estágio de torra (três classes)	CNN MobileNetV2	Exatidão de 97,75%, revocação de 96,44% e precisão de 96,33%	Hakim et al., 2020
Pão	Classificação do estágio de forneamento (sete classes)	Hybrid system CNN-SVM	Redução de 90% no tempo de convergência e exatidão de 99,35% (±0,18%)	Cotrim et al., 2021
Café	Classificação do estágio de torra e predição do tempo de torra (nove classes)	CNN with GAP function	exatidão de 95,83% (±0,32%) e RMSE de 0,4 min (R^2=0,9870)	Cotrim et al., (não publicado)

7. Aprendizado de máquina

A modelagem de processos é particularmente importante para a indústria de alimentos, pois permite que se entenda e conheça com antecedência os resultados de um processo, possibilitando a tomada de decisão e, quando necessário, realização de ações corretivas ou ajustes do mesmo (J. Wang et al., 2018). Na construção de modelos representativos de um processo, a abordagem fenomenológica é preferível, pois possibilita um melhor entendimento dos fatores envolvidos. Entretanto, dada a complexidade dos processos da indústria de alimentos, o uso de técnicas de aprendizado de máquina tem ganhado espaço, pois são facilmente implementáveis e não dependem do conhecimento prévio da relação entre as variáveis de entrada e saída (Trystram, 2012). Foi nesse contexto que um dos primeiros sistemas de visão computacional (SVC) baseado em redes neurais artificiais (ANN) para predição do escurecimento da crosta de biscoitos durante o forneamento foi proposto (Broyart & Trystram, 2003). No estudo, foi adotada uma rede neural artificial recorrente (RNN), onde os dados de entrada foram a cor (L*) da crosta, a umidade e a temperatura no centro da amostra, medidas no tempo t. A RNN foi treinada para estimar a cor da crosta (L*) da amostra no tempo t+1 e apresentou RMSE de 9,39 na etapa de validação do modelo.

Em outra aplicação, máquinas de vetores de suporte (SVM) foram utilizadas para classificar amostras de biscoito quanto ao grau de forneamento, baseado em imagens coletadas ao longo do processo (Nashat & Abdullah, 2010). Neste estudo, diferentes abordagens multiclasses das SVM foram utilizadas para classificar imagens de oito estágios de forneamento de biscoitos. Os melhores resultados foram obtidos pelo método directed acyclic graph (DAG), o qual apresentou exatidão de 87,25%, e o método balanced binary tree (BBT), com exatidão de 86,75%. Os autores argumentam que esse tipo de modelo possui a vantagem de ser facilmente modificado para funcionar com outros tipos de produtos forneados ou torrados.

Na modelagem de processos complexos, como a torra e o forneamento, a escolha das variáveis de entrada é crucial para o sucesso do modelo e deve levar em conta a existência de sensores adequados para a medição das variáveis de entrada e propriedades do produto (Defraeye et al., 2021; Verboven et al., 2020). Com isso em mente, foram utilizadas ANN para construção de modelos preditivos de

cor (L*, a*) da superfície de grãos de café em função dos compostos de aroma liberados pelos grãos durante a torra. Os compostos de aroma foram medidos por um nariz eletrônico composto por 10 sensores semicondutores de óxido metálico e expressos pela condutividade relativa (G/G_0) (Romani et al., 2012). No estudo foram utilizadas redes do tipo multilyer perceptron (MLP) e general regression neural network (GRNN). O modelo GRNN apresentou melhor desempenho geral, com coeficiente de correlação superior a 0,98.

Em outro trabalho, um sistema de monitoramento de cor foi utilizado para determinar a umidade no grão de café ao longo do processo de torra utilizando um modelo Adaptive Network-based Fuzzy Inference System (ANFIS), o qual combina a habilidade de aprendizagem e conectividade das ANN com a lógica humana presente nos sistemas Fuzzy (Virgen-Navarro et al., 2016). No trabalho em questão, imagens do processo de torra foram capturadas em um dispositivo com controle de iluminação. Os valores médios dos pixels presentes na imagem, no espaço de cor RGB, foram convertidos para o espaço de cor CIELab e os valores de L*, a* e b* utilizados para a construção do sistema. O modelo resultante (R^2 > 0,98 e RMSE < 0,002) superou as técnicas convencionais de modelagem do escurecimento do grão de café baseados no comportamento cinético das principais reações químicas envolvidas.

Mais recentemente, foi construído um CVS para estimar o índice de torra do café, na escala SCAA/Agtron, baseado em imagens do café obtidas ao longo do tempo de torra, utilizando uma ANN (Leme et al., 2019). O modelo foi construído em duas etapas. Na primeira, os valores médios de cor dos pixels da imagem (RGB) foram convertidos para os índices de cor L*,a* e b*, no espaço de cor CIELab, mediante uso de uma ANN. Na segunda etapa, os respectivos índices de cor foram utilizados para treinar uma segunda ANN para estimar o índice de torra na escala SCAA/Agtron. O SVC resultante apresentou boa capacidade preditiva (R^2 = 0.99) da cor do café e foi considerado pelos autores como uma ferramenta confiável para automação do processo na indústria.

8. Aprendizado profundo

Embora os modelos de *machine learning* tradicionais tenham alcançados resultados bastante promissores, os mesmos ainda dependem da intervenção

humana na seleção das características que compõe o banco de dados de entrada. Por outro lado, as técnicas de aprendizado profundo, em especial as redes neurais convolucionais (CNN), apresentam como vantagem a sua capacidade de extrair os mapas de características diretamente do banco de dados bruto, sem a necessidade de intervenção humana (J. Wang et al., 2018). Tal vantagem possibilitou um grande salto no desenvolvimento dos sistemas de visão computacional (CVS) observado na última década (Zhu et al., 2021).

Nesse contexto, a modelagem do escurecimento em alimentos, decorrente do processamento térmico, também tem apresentado boas perspectivas. Até pouco tempo se sabia que as CNN eram capazes de extrair apenas características de textura e forma (Gu et al., 2018). Entretanto, recentemente foi demonstrado que as CNN também são capazes de extrair características de cores (Rafegas & Vanrell, 2018). Tal descoberta abriu portas para o desenvolvimento de modelos capazes de lidar com a diversidade de nuances de cores decorrentes do processamento térmico de alimentos. O primeiro trabalho nessa área foi proposto para geração de imagens fotorrealísticas da crosta de pães (Chen et al., 2019). Nesse trabalho, foi utilizada uma Generative adversarial network (GAN), treinada num banco de dados de imagens de crosta de pães com escurecimento induzido por laser, para geração das imagens. Os autores argumentam que o modelo contribui para o desenvolvimento de softwares de desenho assistido por computador (CAD), aplicados ao processamento de alimentos, ao criar modelos fotorrealistas mais precisos.

No ano seguinte, foi proposto um modelo de rede neural convencional com reduzido número de camadas convolucionais (Short-CNN) para acompanhamento e classificação dos estágios de forneamento de pães, baseado unicamente em fragmentos de imagens da crosta do pães (Cotrim et al., 2020). Foi demonstrada que as CNN também são capazes de lidar com as mudanças de cores presentes no escurecimento decorrente do forneamento. A classificação dos estágios de forneamento foi realizada mediante características de cores extraídas diretamente dos fragmentos de imagens da crosta dos pães, com exatidão global de 98,8%. Além disso, o modelo com reduzido número de camadas convolucionais (Short-CNN) mostrou-se mais adequado para lidar com problemas envolvendo mudanças de cores em relação às arquiteturas tradicionais (AlexNet e VGGNet).

Numa outra abordagem, um sistema híbrido foi proposto, se utilizando da capacidade de extrair características de cores das redes neurais convolucionais (CNN) e a capacidade de classificação das máquinas de vetores de suporte (SVM) para classificação dos estágios de forneamento de pães (Cotrim et al., 2021). O sistema híbrido CNN-SVM foi capaz de classificar adequadamente os sete estágios de forneamento utilizados no estudo, com exatidão global média de 99.35% (±0.18%) e redução de 90% no tempo de convergência durante a etapa de treinamento.

A primeira arquitetura empregada na modelagem do escurecimento de grãos de café decorrente da torra foi a MobileNetV2, modelo otimizado para funcionar em sistemas móveis como é o caso de celulares e tabletes. O modelo em questão foi desenvolvido para identificar três graus de torra de café em imagens dos grãos torrados e apresentou exatidão média de 97,75% (Hakim et al., 2020).

Mais recentemente, um modelo de rede neural convolucional com reduzido número de camadas convolucionais, combinado com a função Global Average Pooling (GAP), foi utilizado para modelar um sistema de controle de torra de café assistido por computador (CAS-CNN) (Dados não publicados). O CAS-CNN foi treinado para classificar nove estágios de torra de café e estimar o tempo de torra restante para se atingir o final do processo. No modo de classificação, o CAS-CNN apresentou exatidão global média de 95,83% (±0,32%) e no modo de regressão, para estimativa do tempo restante para o final do processo de torra, o modelo apresentou RMSE de 0,4 min (R^2=0,9870).

9. Desafios e futuros trabalhos

Nas seções anteriores foram discutidas as bases da inteligência artificial e suas aplicações na modelagem do escurecimento dos alimentos decorrentes dos processos de torra e forneamento. Embora aplicações no campo da aprendizagem de máquina e aprendizagem profundo tenham apresentado relativo sucesso na modelagem das mudanças de cor durante a torra ou forneamento, algumas limitações ainda persistem. Os principais desafios a serem superados para uma adoção mais robusta da inteligência artificial podem ser resumidos, conforme se segue:

1. Tamanho e qualidade do banco de dados. Uma vez que as técnicas de inteligência artificial citadas consistem em modelos orientados por dados (Defraeye et al., 2021), é razoável admitir que o banco de dados exerça papel crucial no sucesso da modelagem dos processos. Embora na pesquisa acadêmica o tamanho global do banco de dados represente uma limitação na aplicação dessas técnicas, na indústria o problema do tamanho do banco de dados se apresenta na forma de classes desbalanceadas (Y. Wu et al., 2019). Isso ocorre porque em um processo controlado são raros os casos de não conformidades, o que limita o tamanho de tal classe frente ao número de exemplares dentro das especificações. Além disso, a qualidade do banco de dados também importa e a presença de ruídos ou outliers pode comprometer a qualidade do modelo final, limitando seu uso no mundo real (Dornaika et al., 2020). Assim, é necessária a adoção de estratégias para remoção de ruídos e outliers do banco de dados antes do seu uso no treinamento de um modelo. Para os modelos baseados em imagens, um problema adicional diz respeito aos fatores ambientais, especialmente iluminação. As cores numa imagem capturada por sensores CCD são fortemente influenciadas pelo tipo e intensidade de iluminação (Bianco et al., 2017). Dessa forma, a confiabilidade do modelo passa pelo controle rígido do sistema de iluminação durante o processo de captura de imagens.

2. Seleção do modelo. Como visto, tanto técnicas tradicionais de aprendizado de máquina quanto técnicas de aprendizado profundo são adequadas para a modelagem do escurecimento durante a torra ou forneamento. Entretanto, a escolha do tipo de modelo a ser adotado depende do tipo de banco de dados e dos recursos computacionais disponíveis. Em geral, técnicas de aprendizado de máquina lidam bem com dados de entrada em formatos mais simples e requerem computadores com menor poder computacional. Por outro lado, técnicas de aprendizado profundo se adéquam melhor a banco de dados mais complexos, porém com exigência de computadores com elevada capacidade de processamento. Além disso, dentro do aprendizado profundo, sempre foi consenso que modelos mais profundos apresentavam maior capacidade discriminativa, o que é verdade para problemas com grande número de classes. Porém, tais modelos requerem a adoção de bancos de

dados gigantescos. Entretanto, alguns autores vêm demonstrando que para problemas com pequeno número de classes, e com bancos de dados limitados, a adoção de modelos pouco profundos parece ser uma opção vantajosa (Chen et al., 2019; Cotrim et al., 2020; Q. Zhang et al., 2018). Outro ponto importante diz respeito a adequada seleção de métricas de avaliação dos modelos. Dependendo do tipo de problema, ou mesmo tipo de banco de dados, métricas comuns como Exatidão, *Root Mean Square Error* ou Coeficiente de Determinação podem não ser as mais adequadas (Kotsiopoulos et al., 2021). Dessa forma, é necessário empregar tempo na seleção da métrica mais adequada para avaliação da qualidade do modelo.

3. Aprendizagem incremental. Os modelos de inteligência artificial não foram projetados para aprender de forma incremental (J. Wang et al., 2018). Isso representa um grande problema, especialmente no contexto industrial, onde pequenas mudanças de cenário são comuns, seja pela atualização de um equipamento, modificação numa formulação, etc. Dessa forma, em muitos casos, tais mudanças de cenário resultam na necessidade de treinamento do modelo desde o início. Isso é particularmente dispendioso nos modelos de aprendizado profundo, os quais demanda alto poder computacional na sua implementação. Nesse sentido, faz-se necessário explorar técnicas e modelos que possibilitem o aprendizado incremental, seja mediante técnicas de Transferência de Aprendizagem ou ainda Destilação de Conhecimento (*knowledge distilling*) (Y. Wu et al., 2019; Xu et al., 2019).

Além das oportunidades decorrentes das limitações anteriormente apresentadas, outras surgem à medida que outras técnicas são introduzidas. Nesse contexto, para o desenvolvimento de sistemas de visão computacional, modelos de 3D-CNN utilizando imagens hiperespectrais parecem indicar a direção a se seguir (Y. Liu et al., 2021). Além disso, a combinação da modelagem fenomenológica e em inteligência artificial para a criação de gêmeos digitais parece ser a tendência (Verboven et al., 2020), o que aponta para a necessidade de novos estudos no desenvolvimento de interfaces entre os modelos mecanísticos e de inteligência artificial. Um possível caminho consiste na utilização de modelos de inteligência artificial para classificar ou estimar propriedades dos alimentos a partir dos dados de saída de modelos mecanísticos (Ritto & Rochinha, 2021).

10. Conclusões

A modelagem do escurecimento durante a torra ou forneamento ainda permanece um desafio, devido aos diferentes equipamentos e configurações de processos adotados, além da heterogeneidade das matérias-primas. Entretanto, a modelagem utilizando recursos de inteligência artificial tem alcançado bons resultados, indicando ser um dos caminhos a serem seguidos, especialmente para aplicações industriais, no contexto da industria 4.0. Na presente revisão foram introduzidos os principais modelos de aprendizado de máquina tradicionais e o modelo mais recente de aprendizado profundo, as redes neurais convolucionais. Foi detalhado o modo de funcionamento e os principais componentes de uma CNN. Adicionalmente, foram listadas as principais aplicações das técnicas de inteligência artificial na modelagem do processo de torra e forneamento. Finalmente, foram apresentados os principais desafios a serem superados pelos pesquisadores na adoção das técnicas de inteligência artificial para aplicação em problemas no contexto real. Espera-se que essa revisão possa encorajar mais pesquisas na modelagem do escurecimento em alimentos decorrente dos processos de torra e forneamento.

Referências

Ali, I., Alharbi, O. M. L., Alothman, Z. A., Badjah, A. Y., Alwarthan, A., & Basheer, A. A. (2018). Artificial neural network modelling of amido black dye sorption on iron composite nano material: Kinetics and thermodynamics studies. *Journal of Molecular Liquids*, *250*, 1–8. https://doi.org/10.1016/j.molliq.2017.11.163

Ali, J. M., Hussain, M. A., Tade, M. O., & Zhang, J. (2015). Artificial Intelligence techniques applied as estimator in chemical process systems – A literature survey. *Expert Systems with Applications*, *42*(14), 5915–5931. https://doi.org/10.1016/j.eswa.2015.03.023

Arepally, D., Reddy, R. S., Goswami, T. K., & Datta, A. K. (2020). Biscuit baking: A review. *LWT*, *131*(January), 109726. https://doi.org/10.1016/j.lwt.2020.109726

Baggenstoss, J., Poisson, L., Kaegi, R., Perren, R., & Escher, F. (2008). Coffee Roasting and Aroma Formation: Application of Different Time−Temperature Conditions. *Journal of Agricultural and Food Chemistry*, *56*(14), 5836–5846.

https://doi.org/10.1021/jf800327j

Bai, J.-W., Xiao, H.-W., Ma, H.-L., & Zhou, C.-S. (2018). Artificial Neural Network Modeling of Drying Kinetics and Color Changes of Ginkgo Biloba Seeds during Microwave Drying Process. *Journal of Food Quality*, *2018*, 1–8. https://doi.org/10.1155/2018/3278595

Battineni, G., Chintalapudi, N., & Amenta, F. (2019). Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM). *Informatics in Medicine Unlocked*, *16*(May), 100200. https://doi.org/10.1016/j.imu.2019.100200

Bertrand, E., El Boustany, P., Faulds, C. B., & Berdagué, J.-L. (2018). The Maillard Reaction in Food: An Introduction. *Reference Module in Food Science*, 1–10. https://doi.org/10.1016/b978-0-08-100596-5.21459-5

Bianco, S., Cusano, C., Napoletano, P., & Schettini, R. (2017). Improving CNN-Based Texture Classification by Color Balancing. *Journal of Imaging*, *3*(3), 33. https://doi.org/10.3390/jimaging3030033

Bicho, N. C., Leitão, A. E., Ramalho, J. C., & Lidon, F. C. (2012). Use of colour parameters for roasted coffee assessment. *Food Science and Technology*, *32*(3), 436–442. https://doi.org/10.1590/s0101-20612012005000068

Bini, S. A. (2018). Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *Journal of Arthroplasty*, *33*(8), 2358–2361. https://doi.org/10.1016/j.arth.2018.02.067

Bortolini, M., Ferrari, E., Gamberi, M., Pilati, F., & Faccio, M. (2017). Assembly system design in the Industry 4.0 era: a general framework. *IFAC-PapersOnLine*, *50*(1), 5700–5705. https://doi.org/10.1016/j.ifacol.2017.08.1121

Bottazzi, D., Farina, S., Milani, M., & Montorsi, L. (2012). A numerical approach for the analysis of the coffee roasting process. *Journal of Food Engineering*, *112*(3), 243–252. https://doi.org/10.1016/j.jfoodeng.2012.04.009

Broyart, B., & Trystram, G. (2003). Modelling of Heat and Mass Transfer Phenomena

and Quality Changes During Continuous Biscuit Baking Using Both Deductive and Inductive (Neural Network) Modelling Principles. *Food and Bioproducts Processing*, *81*(4), 316–326. https://doi.org/10.1205/096030803322756402

Broyart, B., Trystram, G., & Duquenoy, A. (1998). Predicting colour kinetics during cracker baking. *Journal of Food Engineering*, *35*(3), 351–368. https://doi.org/10.1016/S0260-8774(98)00021-1

Buera, M. del P., Chirife, J., Resnik, S. L., & Lozano, R. D. (1987). Nonenzymatic Browning in Liquid Model Systems of High Water Activity: Kinetics of Color Changes due to Caramelization of Various Single Sugars. *Journal of Food Science*, *52*(4), 1059–1062. https://doi.org/10.1111/j.1365-2621.1987.tb14275.x

Bustos-Vanegas, J. D., Corrêa, P. C., Martins, M. A., Baptestini, F. M., Campos, R. C., de Oliveira, G. H. H., & Nunes, E. H. M. (2018). Developing predictive models for determining physical properties of coffee beans during the roasting process. *Industrial Crops and Products*, *112*(September 2017), 839–845. https://doi.org/10.1016/j.indcrop.2017.12.015

Cao, W., Wang, X., Ming, Z., & Gao, J. (2018). A review on neural networks with random weights. *Neurocomputing*, *275*, 278–287. https://doi.org/10.1016/j.neucom.2017.08.040

Castro, W., Oblitas, J., Chuquizuta, T., & Avila-George, H. (2017). Application of image analysis to optimization of the bread-making process based on the acceptability of the crust color. *Journal of Cereal Science*, *74*, 194–199. https://doi.org/10.1016/j.jcs.2017.02.002

Çelik, E. E., & Gökmen, V. (2020). Formation of Maillard reaction products in bread crust-like model system made of different whole cereal flours. *European Food Research and Technology*, *246*(6), 1207–1218. https://doi.org/10.1007/s00217-020-03481-4

Cha, C.-Y., & Lee, K.-G. (2020). Effect of roasting conditions on the formation and kinetics of furan in various nuts. *Food Chemistry*, *331*(January), 127338. https://doi.org/10.1016/j.foodchem.2020.127338

Chaibakhsh, A., Chaibakhsh, N., & Rahman, M. B. A. (2011). Application of Fuzzy

Modeling and Optimization in Enzymatic Esterification Process. *International Journal of Chemical Engineering and Applications*, *2*(6), 408–415. https://doi.org/10.7763/IJCEA.2011.V2.143

Chen, P. Y., Blutinger, J. D., Meijers, Y., Zheng, C., Grinspun, E., & Lipson, H. (2019). Visual modeling of laser-induced dough browning. *Journal of Food Engineering*, *243*(April 2018), 9–21. https://doi.org/10.1016/j.jfoodeng.2018.08.022

Chhanwal, N., Indrani, D., Raghavarao, K. S. M. S., & Anandharamakrishnan, C. (2011). Computational fluid dynamics modeling of bread baking process. *Food Research International*, *44*(4), 978–983. https://doi.org/10.1016/j.foodres.2011.02.037

Chindapan, N., Soydok, S., & Devahastin, S. (2019). Roasting Kinetics and Chemical Composition Changes of Robusta Coffee Beans During Hot Air and Superheated Steam Roasting. *Journal of Food Science*, *84*(2), 292–302. https://doi.org/10.1111/1750-3841.14422

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

Cotrim, W. da S., Felix, L. B., Minim, V. P. R., Campos, R. C., & Minim, L. A. (2021). Development of a hybrid system based on convolutional neural networks and support vector machines for recognition and tracking color changes in food during thermal processing. *Chemical Engineering Science*, *240*, 116679. https://doi.org/10.1016/j.ces.2021.116679

Cotrim, W. da S., Minim, V. P. R., Felix, L. B., & Minim, L. A. (2020). Short convolutional neural networks applied to the recognition of the browning stages of bread crust. *Journal of Food Engineering*, *277*, 109916. https://doi.org/10.1016/j.jfoodeng.2020.109916

Czech, H., Schepler, C., Klingbeil, S., Ehlert, S., Howell, J., & Zimmermann, R. (2016). Resolving Coffee Roasting-Degree Phases Based on the Analysis of Volatile Compounds in the Roasting Off-Gas by Photoionization Time-of-Flight Mass Spectrometry (PI-TOFMS) and Statistical Data Analysis: Toward a PI-

TOFMS Roasting Model. *Journal of Agricultural and Food Chemistry*, *64*(25), 5223–5231. https://doi.org/10.1021/acs.jafc.6b01683

Debska, B., & Guzowska-Świder, B. (2011). Application of artificial neural network in food classification. *Analytica Chimica Acta*, *705*(1–2), 283–291. https://doi.org/10.1016/j.aca.2011.06.033

Defraeye, T., Shrivastava, C., Berry, T., Verboven, P., Onwude, D., Schudel, S., Bühlmann, A., Cronje, P., & Rossi, R. M. (2021). Digital twins are coming: Will we need them in supply chains of fresh horticultural produce? *Trends in Food Science & Technology*, *109*(March 2020), 245–258. https://doi.org/10.1016/j.tifs.2021.01.025

Delgado, A., Rauh, C., Park, J., Kim, Y., Groß, F., & Diez, L. (2016). Artificial Neural Networks: Applications in Food Processing. In *Reference Module in Food Science*. Elsevier. https://doi.org/10.1016/b978-0-08-100596-5.03125-5

Demir, A. D., Celayeta, J. M. F., Cronin, K., & Abodayeh, K. (2002). Modelling of the kinetics of colour change in hazelnuts during air roasting. *Journal of Food Engineering*, *55*(4), 283–292. https://doi.org/10.1016/S0260-8774(02)00103-6

Dornaika, F., Bekhouche, S., & Arganda-Carreras, I. (2020). Robust regression with deep CNNs for facial age estimation: An empirical study. *Expert Systems with Applications*, *141*, 112942. https://doi.org/10.1016/j.eswa.2019.112942

Funes, E., Allouche, Y., Beltrán, G., & Jiménez, A. (2015). A Review: Artificial Neural Networks as Tool for Control Food Industry Process. *Journal of Sensor Technology*, *05*(01), 28–43. https://doi.org/10.4236/jst.2015.51004

Getachew, A. T., & Chun, B.-S. (2018). Coffee Flavor. In *Encyclopedia of Food Chemistry*. Elsevier. https://doi.org/10.1016/b978-0-08-100596-5.21658-2

Gonçalves, E. C., Minim, L. A., Coimbra, J. S. R., & Minim, V. P. R. (2005). Modeling sterilization process of canned foods using artificial neural networks. *Chemical Engineering and Processing: Process Intensification*, *44*(12), 1269–1276. https://doi.org/10.1016/j.cep.2005.04.001

Górska-Horczyczak, E., Horczyczak, M., Guzek, D., Wojtasik-Kalinowska, I., &

Wierzbicka, A. (2017). Chromatographic fingerprints supported by artificial neural network for differentiation of fresh and frozen pork. *Food Control*, *73*, 237–244. https://doi.org/10.1016/j.foodcont.2016.08.010

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, *187*, 27–48. https://doi.org/10.1016/j.neucom.2015.09.116

Hadiyanto, Asselman, A., Straten, G. van, Boom, R. M., Esveld, D. C., & Boxtel, A. J. B. van. (2007). Quality prediction of bakery products in the initial phase of process design. *Innovative Food Science & Emerging Technologies*, *8*(2), 285–298. https://doi.org/10.1016/j.ifset.2007.01.006

Hakim, M., Djatna, T., & Yuliasih, I. (2020). Deep learning for roasting coffee bean quality assessment using computer vision in mobile environment. *2020 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2020*. https://doi.org/10.1109/ICACSIS51025.2020.9263224

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Helou, C., Jacolot, P., Niquet-Léridon, C., Gadonna-Widehem, P., & Tessier, F. J. (2016). Maillard reaction products in bread: A novel semi-quantitative method for evaluating melanoidins in bread. *Food Chemistry*, *190*, 904–911. https://doi.org/10.1016/j.foodchem.2015.06.032

Hernández, J. A., Heyd, B., & Trystram, G. (2008). Prediction of brightness and surface area kinetics during coffee roasting. *Journal of Food Engineering*, *89*(2), 156–163. https://doi.org/10.1016/j.jfoodeng.2008.04.026

Kim, S. Y., Ko, J. A., Kang, B. S., & Park, H. J. (2018). Prediction of key aroma

development in coffees roasted to different degrees by colorimetric sensor array. *Food Chemistry*, *240*(July 2017), 808–816. https://doi.org/10.1016/j.foodchem.2017.07.139

Kocadağli, T., & Gökmen, V. (2016). Multiresponse kinetic modelling of Maillard reaction and caramelisation in a heated glucose/wheat flour system. *Food Chemistry*, *211*, 892–902. https://doi.org/10.1016/j.foodchem.2016.05.150

Kotsiopoulos, T., Sarigiannidis, P., Ioannidis, D., & Tzovaras, D. (2021). Machine Learning and Deep Learning in smart manufacturing: The Smart Grid paradigm. *Computer Science Review*, *40*, 100341. https://doi.org/10.1016/j.cosrev.2020.100341

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, *25*, 1–9. https://doi.org/10.1016/B978-008046518-0.00119-7

Laguna, L., Vallons, K. J. R., Jurgens, A., & Sanz, T. (2013). Understanding the Effect of Sugar and Sugar Replacement in Short Dough Biscuits. *Food and Bioprocess Technology*, *6*(11), 3143–3154. https://doi.org/10.1007/s11947-012-0968-5

Le Cun, Y., Jackel, L. D., Boser, B., Denker, J. S., Graf, H. P., Guyon, I., Henderson, D., Howard, R. E., & Hubbard, W. (1989). Handwritten digit recognition: applications of neural network chips and automatic learning. *IEEE Communications Magazine*, *27*(11), 41–46. https://doi.org/10.1109/35.41400

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

Lee, J., Davari, H., Singh, J., & Pandhare, V. (2018). Industrial Artificial Intelligence for industry 4.0-based manufacturing systems. *Manufacturing Letters*, *18*(September), 20–23. https://doi.org/10.1016/j.mfglet.2018.09.002

Leme, D. S., da Silva, S. A., Barbosa, B. H. G., Borém, F. M., & Pereira, R. G. F. A. (2019). Recognition of coffee roasting degree using a computer vision system.

Computers and Electronics in Agriculture, *156*(October 2018), 312–317. https://doi.org/10.1016/j.compag.2018.11.029

León-Roque, N., Abderrahim, M., Nuñez-Alejos, L., Arribas, S. M., & Condezo-Hoyos, L. (2016). Prediction of fermentation index of cocoa beans (Theobroma cacao L.) based on color measurement and artificial neural networks. *Talanta*, *161*, 31–39. https://doi.org/10.1016/j.talanta.2016.08.022

Li, H. X., & Si, H. (2017). Control for Intelligent Manufacturing: A Multiscale Challenge. *Engineering*, *3*(5), 608–615. https://doi.org/10.1016/J.ENG.2017.05.016

Liu, N., Han, J., Zhang, D., Shifeng Wen, & Liu, T. (2015). Predicting eye fixations using convolutional neural networks. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 362–370. https://doi.org/10.1109/CVPR.2015.7298633

Liu, Y., Pu, H., & Sun, D. (2021). Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices. *Trends in Food Science & Technology*, *113*(May), 193–204. https://doi.org/10.1016/j.tifs.2021.04.042

McCall, J. (2005). Genetic algorithms for modelling and optimisation. *Journal of Computational and Applied Mathematics*, *184*(1), 205–222. https://doi.org/10.1016/j.cam.2004.07.034

Michele, A., Colin, V., & Santika, D. D. (2019). Mobilenet convolutional neural networks and support vector machines for palmprint recognition. *Procedia Computer Science*, *157*, 110–117. https://doi.org/10.1016/j.procs.2019.08.147

Muller, K.-R., Mika, S., Ratsch, G., Tsuda, K., & Scholkopf, B. (2001). An introduction to kernel-based learning algorithms. *IEEE Transactions on Neural Networks*, *12*(2), 181–201. https://doi.org/10.1109/72.914517

Mundt, S., & Wedzicha, B. L. (2007). A kinetic model for browning in the baking of biscuits: Effects of water activity and temperature. *LWT - Food Science and Technology*, *40*(6), 1078–1082. https://doi.org/10.1016/j.lwt.2006.07.011

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In J. Fürnkranz & T. Joachims (Eds.), *Proceedings of the 27th International Conference on Machine Learningence* (pp. 807–814). Omnipress. https://icml.cc/Conferences/2010/proceedings.html

Nalepa, J., & Kawulok, M. (2019). Selecting training sets for support vector machines: a review. *Artificial Intelligence Review*, *52*(2), 857–900. https://doi.org/10.1007/s10462-017-9611-1

Nashat, S., & Abdullah, M. Z. (2010). Multi-class colour inspection of baked foods featuring support vector machine and Wilk's λ analysis. *Journal of Food Engineering*, *101*(4), 370–380. https://doi.org/10.1016/j.jfoodeng.2010.07.022

Patan, K. (2018). Two stage neural network modelling for robust model predictive control. *ISA Transactions*, *72*, 56–65. https://doi.org/10.1016/j.isatra.2017.10.011

Pour-Damanab, A. S., Jafary, A., & Rafiee, S. (2014). Kinetics of the crust thickness development of bread during baking. *Journal of Food Science and Technology*, *51*(11), 3439–3445. https://doi.org/10.1007/s13197-012-0872-z

Purlis, E. (2011). Bread baking: Technological considerations based on process modelling and simulation. *Journal of Food Engineering*, *103*(1), 92–102. https://doi.org/10.1016/j.jfoodeng.2010.10.003

Purlis, E. (2012). Baking process design based on modelling and simulation: Towards optimization of bread baking. *Food Control*, *27*(1), 45–52. https://doi.org/10.1016/j.foodcont.2012.02.034

Purlis, E., & Salvadori, V. O. (2007). Bread browning kinetics during baking. *Journal of Food Engineering*, *80*(4), 1107–1115. https://doi.org/10.1016/j.jfoodeng.2006.09.007

Purlis, E., & Salvadori, V. O. (2009). Modelling the browning of bread during baking. *Food Research International*, *42*(7), 865–870. https://doi.org/10.1016/j.foodres.2009.03.007

Putranto, A., Chen, X. D., Devahastin, S., Xiao, Z., & Webley, P. A. (2011). Application of the reaction engineering approach (REA) for modeling intermittent

drying under time-varying humidity and temperature. *Chemical Engineering Science*, *66*(10), 2149–2156. https://doi.org/10.1016/j.ces.2011.02.025

Putranto, A., Chen, X. D., & Zhou, W. (2015). Bread baking and its color kinetics modeled by the spatial reaction engineering approach (S-REA). *Food Research International*, *71*, 58–67. https://doi.org/10.1016/j.foodres.2015.01.029

Quintas, M., Guimarães, C., Baylina, J., Brandão, T. R. S., & Silva, C. L. M. (2007). Multiresponse modelling of the caramelisation reaction. *Innovative Food Science and Emerging Technologies*, *8*(2), 306–315. https://doi.org/10.1016/j.ifset.2007.02.002

Rafegas, I., & Vanrell, M. (2018). Color encoding in biologically-inspired convolutional neural networks. *Vision Research*, *151*(May), 7–17. https://doi.org/10.1016/j.visres.2018.03.010

Ramírez-Jiménez, A., Guerra-Hernández, E., & García-Villanova, B. (2000). Browning Indicators in Bread. *Journal of Agricultural and Food Chemistry*, *48*(9), 4176–4181. https://doi.org/10.1021/jf9907687

Razzaghi, T., Roderick, O., Safro, I., & Marko, N. (2016). Multilevel weighted support vector machine for classification on healthcare data with missing values. *PLoS ONE*, *11*(5), 1–18. https://doi.org/10.1371/journal.pone.0155119

Ritto, T. G., & Rochinha, F. A. (2021). Digital twin, physics-based model, and machine learning applied to damage detection in structures. *Mechanical Systems and Signal Processing*, *155*, 107614. https://doi.org/10.1016/j.ymssp.2021.107614

Romani, S., Cevoli, C., Fabbri, A., Alessandrini, L., & Dalla Rosa, M. (2012). Evaluation of Coffee Roasting Degree by Using Electronic Nose and Artificial Neural Network for Off-line Quality Control. *Journal of Food Science*, *77*(9), C960–C965. https://doi.org/10.1111/j.1750-3841.2012.02851.x

Sacchetti, G., Di Mattia, C., Pittia, P., & Mastrocola, D. (2009). Effect of roasting degree, equivalent thermal effect and coffee type on the radical scavenging activity of coffee brews and their phenolic fraction. *Journal of Food Engineering*, *90*(1), 74–80. https://doi.org/10.1016/j.jfoodeng.2008.06.005

Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Shibukawa, S., Sugiyama, K., & Yano, T. (1989). Effects of Heat Transfer by Radiation and Convection on Browning of Cookies at Baking. *Journal of Food Science*, *54*(3), 621–624. https://doi.org/10.1111/j.1365-2621.1989.tb04666.x

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. In Y. Bengio & Y. LeCun (Eds.), *3rd International Conference on Learning Representations, ICLR 2015* (pp. 1–14). http://arxiv.org/abs/1409.1556

Sruthi, N. U., Premjit, Y., Pandiselvam, R., Kothakota, A., & Ramesh, S. V. (2021). An overview of conventional and emerging techniques of roasting: Effect on food bioactive signatures. *Food Chemistry*, *348*(October 2020), 129088. https://doi.org/10.1016/j.foodchem.2021.129088

Stamp, J. A., & Labuza, T. P. (1983). Kinetics of the Maillard Reaction Between Aspartame and Glucose in Solution at High Temperatures. *Journal of Food Science*, *48*(2), 543–544. https://doi.org/10.1111/j.1365-2621.1983.tb10785.x

Sudha, L., Dillibabu, R., Srivatsa Srinivas, S., & Annamalai, A. (2016). Optimization of process parameters in feed manufacturing using artificial neural network. *Computers and Electronics in Agriculture*, *120*, 1–6. https://doi.org/10.1016/j.compag.2015.11.004

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *07-12-June*, 1–9. https://doi.org/10.1109/CVPR.2015.7298594

Toker, O. S., Palabiyik, I., Pirouzian, H. R., Aktar, T., & Konar, N. (2020). Chocolate aroma: Factors, importance and analysis. *Trends in Food Science & Technology*, *99*(January), 580–592. https://doi.org/10.1016/j.tifs.2020.03.035

Tomar, D., & Agarwal, S. (2015). Twin Support Vector Machine: A review from 2007 to 2014. *Egyptian Informatics Journal*, *16*(1), 55–69.

https://doi.org/10.1016/j.eij.2014.12.003

Trystram, G. (2012). Modelling of food and food processes. *Journal of Food Engineering*, *110*(2), 269–277. https://doi.org/10.1016/j.jfoodeng.2011.05.001

Tseng, T.-L. (Bill), Aleti, K. R., Hu, Z., & Kwon, Y. (James). (2016). E-quality control: A support vector machines approach. *Journal of Computational Design and Engineering*, *3*(2), 91–101. https://doi.org/10.1016/j.jcde.2015.06.010

Unnikrishnan, A., V, S., & K P, S. (2018). Deep AlexNet with Reduced Number of Trainable Parameters for Satellite Image Classification. *Procedia Computer Science*, *143*, 931–938. https://doi.org/10.1016/j.procs.2018.10.342

Vanin, F. M., Lucas, T., & Trystram, G. (2009). Crust formation and its role during bread baking. *Trends in Food Science and Technology*, *20*(8), 333–343. https://doi.org/10.1016/j.tifs.2009.04.001

Vargas-Elías, G. A., Corrêa, P. C., Souza, N. R. de, Baptestini, F. M., & Melo, E. de C. (2016). Kinetics of mass loss of arabica coffee during roasting process. *Engenharia Agrícola*, *36*(2), 300–308. https://doi.org/10.1590/1809-4430-Eng.Agric.v36n2p300-308/2016

Varol, T., Canakci, A., & Ozsahin, S. (2018). Prediction of effect of reinforcement content, flake size and flake time on the density and hardness of flake AA2024-SiC nanocomposites using neural networks. *Journal of Alloys and Compounds*, *739*, 1005–1014. https://doi.org/10.1016/j.jallcom.2017.12.256

Verboven, P., Defraeye, T., Datta, A. K., & Nicolai, B. (2020). Digital twins of food process operations: the next step for food process models? *Current Opinion in Food Science*, *35*, 79–87. https://doi.org/10.1016/j.cofs.2020.03.002

Virgen-Navarro, L., Herrera-López, E. J., Corona-González, R. I., Arriola-Guevara, E., & Guatemala-Morales, G. M. (2016). Neuro-fuzzy model based on digital images for the monitoring of coffee bean color during roasting in a spouted bed. *Expert Systems with Applications*, *54*, 162–169. https://doi.org/10.1016/j.eswa.2016.01.027

Wählby, U., & Skjöldebrand, C. (2002). Reheating characteristics of crust formed on

buns, and crust formation. *Journal of Food Engineering*, *53*(2), 177–184. https://doi.org/10.1016/S0260-8774(01)00154-6

Wang, J., Ma, Y., Zhang, L., Gao, R. X., & Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems*, *48*, 144–156. https://doi.org/10.1016/j.jmsy.2018.01.003

Wang, X., & Lim, L.-T. (2014a). A Kinetics and Modeling Study of Coffee Roasting Under Isothermal Conditions. *Food and Bioprocess Technology*, *7*(3), 621–632. https://doi.org/10.1007/s11947-013-1159-8

Wang, X., & Lim, L. T. (2014b). Effect of roasting conditions on carbon dioxide degassing behavior in coffee. *Food Research International*, *61*, 144–151. https://doi.org/10.1016/j.foodres.2014.01.027

Wang, X., & Lim, L. T. (2017). Investigation of CO2 precursors in roasted coffee. *Food Chemistry*, *219*, 185–192. https://doi.org/10.1016/j.foodchem.2016.09.095

Wani, I. A., Hamid, H., Hamdani, A. M., Gani, A., & Ashwar, B. A. (2017). Physico-chemical, rheological and antioxidant properties of sweet chestnut (Castanea sativa Mill.) as affected by pan and microwave roasting. *Journal of Advanced Research*, *8*(4), 399–405. https://doi.org/10.1016/j.jare.2017.05.005

Wei, H. M., Su, G. H., Tian, W. X., Qiu, S. Z., & Ni, W. F. (2009). Study on the characteristic points of boiling curve by using wavelet analysis and genetic neural network. *Nuclear Engineering and Design*, *239*(11), 2317–2325. https://doi.org/10.1016/j.nucengdes.2009.07.016

Wei, N. C., Hussain, M. A., & Wahab, A. K. A. (2008). Control of a Batch Polymerization System Using Hybrid Neural Network - First Principle Model. *The Canadian Journal of Chemical Engineering*, *85*(6), 936–945. https://doi.org/10.1002/cjce.5450850616

Willis, M. J., Di Massimo, C., Montague, G. A., Tham, M. T., & Morris, A. J. (1991). Artificial neural networks in process engineering. *IEE Proceedings D Control Theory and Applications*, *138*(3), 256. https://doi.org/10.1049/ip-d.1991.0036

Wu, H., & Zhao, J. (2018). Deep convolutional neural network model based chemical

process fault diagnosis. *Computers and Chemical Engineering*, *115*, 185–197. https://doi.org/10.1016/j.compchemeng.2018.04.009

Wu, Y., Chen, Y., Wang, L., Ye, Y., Liu, Z., Guo, Y., & Fu, Y. (2019). Large Scale Incremental Learning. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, *2019-June*, 374–382. https://doi.org/10.1109/CVPR.2019.00046

Xu, Y., Sun, Y., Liu, X., & Zheng, Y. (2019). A Digital-Twin-Assisted Fault Diagnosis Using Deep Transfer Learning. *IEEE Access*, *7*, 19990–19999. https://doi.org/10.1109/ACCESS.2018.2890566

Zanoni, B., Peri, C., & Bruno, D. (1995). Modelling of browning kinetics of bread crust during baking. *LWT - Food Science and Technology*, *28*(6), 604–609. https://doi.org/10.1016/0023-6438(95)90008-X

Zhang, L., Putranto, A., Zhou, W., Boom, R. M., Schutyser, M. A. I., & Chen, X. D. (2016). Miniature bread baking as a timesaving research approach and mathematical modeling of browning kinetics. *Food and Bioproducts Processing*, *100*, 401–411. https://doi.org/10.1016/j.fbp.2016.08.007

Zhang, Q., Zhuo, L., Li, J., Zhang, J., Zhang, H., & Li, X. (2018). Vehicle color recognition using Multiple-Layer Feature Representations of lightweight convolutional neural network. *Signal Processing*, *147*, 146–153. https://doi.org/10.1016/j.sigpro.2018.01.021

Zhong, R. Y., Xu, X., Klotz, E., & Newman, S. T. (2017). Intelligent Manufacturing in the Context of Industry 4.0: A Review. *Engineering*, *3*(5), 616–630. https://doi.org/10.1016/J.ENG.2017.05.015

Zhu, L., Spachos, P., Pensini, E., & Plataniotis, K. N. (2021). Deep learning and machine vision for food processing: A survey. *Current Research in Food Science*, *4*(March), 233–249. https://doi.org/10.1016/j.crfs.2021.03.009

3. ARTIGO 2 - SHORT CONVOLUTIONAL NEURAL NETWORKS APPLIED TO THE RECOGNITION OF THE BROWNING STAGES OF BREAD CRUST

Neste artigo será introduzida uma rede neural convolucional com reduzido número de camadas convolucionais (Short-CNN) para classificação dos estágios de forneamento de pães, baseado unicamente em fragmentos de imagens da crosta de pães. Será demonstrado que a Short-CNN supera a performance de arquiteturas tradicionais, embora apresente menor requerimento de recursos computacionais. Também será demonstrado que a Short-CNN é capaz de extrair características de cor das imagens e utilizá-las para classificação dos estágios de forneamento das amostras. A seguir é reproduzido o inteiro teor do manuscrito publicado no *Journal of Food Engineering*.

Cotrim, W. da S., Minim, V. P. R., Felix, L. B., & Minim, L. A. (2020). Short convolutional neural networks applied to the recognition of the browning stages of bread crust. **Journal of Food Engineering**, 277, 109916. https://doi.org/10.1016/j.jfoodeng.2020.109916

SHORT CONVOLUTIONAL NEURAL NETWORKS APPLIED TO THE RECOGNITION OF THE BROWNING STAGES OF BREAD CRUST

1 Weskley da Silva Cotrim[a], Valéria Paula Rodrigues Minim[b], Leonardo Bonato Felix[c],
2 Luis Antonio Minim[d,*]

[a] Universidade Federal de Mato Grosso, *Campus* Universitário do Araguaia, Instituto de Ciências Exatas e da Terra, Avenida Valdon Varjão, nº 6.390. Barra do Garças - Mato Grosso, Brasil, CEP: 78600-000, weskleycotrim@ufmt.br
[b] Universidade Federal de Viçosa, Departamento de Tecnologia de Alimentos, *Campus* Universitário – Viçosa – Minas Gerais, Brasil, CEP: 36570-000, vprm@ufv.br
[c] Universidade Federal de Viçosa, Departamento de Engenharia Elétrica, *Campus* Universitário – Viçosa – Minas Gerais, Brasil, CEP: 36570-000, leobonato@ufv.br
[d] Universidade Federal de Viçosa, Departamento de Tecnologia de Alimentos, *Campus* Universitário – Viçosa – Minas Gerais, Brasil, CEP: 36570-000, lminim@ufv.br

*Corresponding author
Prof. Luis Antônio Minim
Universidade Federal de Viçosa, Department of Food Technology, *Campus* Universitário – Viçosa – Minas Gerais, Brasil, CEP: 36570-000, lminim@ufv.br

Abbreviations
CIELab - *Commission internationale de l'éclairage* color space
CL - Convolutional Layer
CNN - Convolutional Neural Network
CVS - Computational Vision System
FC - Fully Connected Region
IC - Inception Module
MLP - MultiLayer Perceptron
RGB – Red, Green, Blue color space
sRGB - Standard RGB
VGGNet - Visual Geometry Group Network

Conflict of interest
The authors have declared no conflicts of interest for this article.

ABSTRACT
A Computational Vision System (CVS) based on a Convolutional Neural Network (CNN), operating with the module Inception v3 and reduced number of convolutional layers (Short-CNN), was proposed for the classification of browning degree of bread crust during baking. The training (70%), validation (15%), and testing (15%) of the CNN was performed using a dataset composed of 374 bread crust image fragments (600 x 600 pixels) over seven baking periods. The resulting CVS does not depend on process variables, overcoming a limitation present in ordinary models. In addition, the CVS was able to correctly classify the images from the test dataset uniformly, being

able to extract the main colors present in the images from the dataset already in the first convolutional layer. The results showed a potential use of CNN in the food industry process control system involving color changes.

Keywords: Industry 4.0, Deep Learning, Process Control, Browning, Baking.

1. INTRODUCTION

During baking, bread undergoes a series of transformations in its texture and color due to moisture loss through evaporation and browning on the surface due to Maillard and Caramelization reactions (Zhang et al., 2016). Such reactions start when the surface temperature reaches values from 105 °C to 115 °C, which occurs when a low moisture crust is formed. Under these conditions, colored compounds (melanoidins) are produced from carbohydrates and there is a release of compounds responsible for the characteristic aroma of the baked product (Vanin et al., 2009).

Changes in color are especially important because they are an indicator of the quality of the process and a fundamental attribute in the acceptance of the product, being closely linked to consumer purchase decisions (Castro et al., 2017; Purlis, 2012). Traditionally, color measurements can be made indirectly by determining the content of melanoidins present in the crust of the bread by physical-chemical analysis (Fogliano and Morales, 2011), spectrophotometric techniques (Helou et al., 2016) and chromatographic methods (Kocadağli and Gökmen, 2016). Color changes can be also measured directly by instrumental colorimetric analysis (Purlis, 2012) or by sensory analysis techniques (Castro et al., 2017).

These methods have, as a main characteristic, a high implementation cost or long time necessary to obtain the results. Although they are very precise, these methods are inadequate for use in industrial control systems, especially in the context of the 4th generation industry (Industry 4.0). Industry 4.0 recommends the use of intelligent equipment and systems, in which the information should be easily obtained and shared autonomously, allowing rapid decision making, with minimization of losses and costs (Li and Si, 2017; Wang et al., 2018; Zhong et al., 2017).

In this context, cheap and fast methods for color measurement and tracking during the baking have been pursued. To satisfy Industry 4.0 assumptions, a proposed technique should be based on objective and non-destructive measures that can be applied to the processing line. In this context, the Computational Vision Systems (CVS) appears as a strategic technique in color analysis. CVS simulates the mode of capturing and interpreting information by the human eye, allowing the objective measurement of color by correlating it with consumer acceptance characteristics or product parameters (Castro et al., 2017).

One of the first CVS proposed to measure the degree of browning of bread during baking was based on evidence that the Maillard and Caramelization reactions follow first-order kinetics. In this model, the color index for each baking stage was

based on the three components (L*, a*, b*) of the CIELab color space and the reaction constant, due to temperature dependence, in the Arrhenius equation. The resulting model presented behavior compatible with the experimental data. However, its use is limited due to the large deviations observed for some oven temperatures (Zanoni et al., 1995) and the high cost of sensors operating in the CIELab color space (Schmittmann and Lammers, 2017).

To overcome these limitations, new models were proposed adopting color spaces based on low-cost sensors, such as the RGB (Red, Blue, Green) space present in cameras. They also included in the model the sample surface temperature and water activity of the bread crust (Purlis and Salvadori, 2009) or the sample surface temperature and the moisture of the bread crust in the estimation of the reaction constant (Putranto et al., 2015; Zhang et al., 2016). However, these variables are known to be difficult to measure experimentally, and their estimates via auxiliary models are necessary (Purlis and Salvadori, 2009). Thus, although the CVS has shown an improvement in the ability to predict the browning of the bread crust, large deviations from the experimental values were still observed. These deviations were attributed to the presence of nonlinearities of the browning process and the need to use supplementary models to predict the values of temperature, water activity or crust moisture (Purlis and Salvadori, 2009; Putranto et al., 2015).

In parallel, a CVS based on Neural Networks was proposed (Broyart and Trystram, 2003), due to its ability to deal with nonlinearities present in the common processes of the food industry and its learning capacity based on a set of previous data from the process, dispensing the adoption of mechanistic models (Varol et al., 2018). However, in this type of CVS additional variables were still maintained, such as sample temperature and crust moisture (Broyart and Trystram, 2003). The resulting CVS showed good results, although the observed deviations point to the need to use a model that does not depend on other variables, besides those directly related to the color of the sample.

Thus, a possible solution would be the adoption of Convolutional Neural Networks (CNN), since they have stood out for their ability to recognize complex patterns present in images (He et al., 2016; Krizhevsky et al., 2012; Simonyan and Zisserman, 2015). CNN have been widely applied in the detection and classification of objects in images based on their texture and shape. In industrial processes, the CNN have already been applied in the detection of the presence of insects in stored grains (Shen et al., 2018), in the diagnosis and detection of failures in the chemical industry (Wu and Zhao, 2018) and in the detection of small impurities in bottled beverages (Guo et al., 2018). However, recent studies have demonstrated the potential use of CNN in solving problems involving colors, having been applied in the classification of the maturation stage of pineapple (Azman and Ismail, 2017) and bananas (Y. Zhang et al., 2018) and in the identification of colors in vehicles (Q. Zhang et al., 2018).

In solving color problems, AlexNet architecture is the main CNN used (Aarathi and Abraham, 2017). Originally proposed by Alex Krizhevsky, AlexNet is composed of seven hidden layers, five of which are convolutional and two are fully connected (Krizhevsky et al., 2012). The convolutional layers have the attribute to extract the feature map of the image, storing such information in convolutional kernels. Additional information about training mode and the main equations involved can be found in Gu et al. (2018). Another architecture that has been successfully used in problems involving colors is called VGGNet (Rafegas and Vanrell, 2018). More robust than AlexNet, it can have up to 19 layers, divided into convolutional and fully connected (Simonyan and Zisserman, 2015).

More recently, it was demonstrated that the first five convolutional layers have convolutional kernels involved in the extraction of colors present in the image (Rafegas and Vanrell, 2018). In the same direction, an architecture with a reduced number of convolutional layers for color identification was proposed (Q. Zhang et al., 2018). This network has three convolutional layers, followed by a Global Average Pooling layer, which connects directly to the classification layer. The authors argue that the use of networks with a reduced number of convolutional layers brings two direct benefits: i) reduction in the demand for computational processing in the training stage and ii) lower storage space requirement of the network parameters already trained, being an objective to be pursued.

Another important aspect in defining the architecture of convolutional neural networks used to solve problems involving colors is the size of convolutional kernels. Although there is no consensus on the ideal kernel size, good results were obtained in convolutional networks with 11x11, 5x5 and 3x3 kernel sizes (Aarathi and Abraham, 2017), 7x7, 5x5 and 3x3 (Rafegas and Vanrell, 2018) or 7x7 and 3x3 (Q. Zhang et al., 2018). Such results indicate that a network capable of choosing the kernel size can bring benefits to build networks with reduced number of convolutional layers. In this sense, the Inception v3 module seems to be very suitable for solving this type of problem, since it allows different convolutional kernel sizes to be applied to the same layer (Szegedy et al., 2015). The Inception v3 module is composed of five convolutional layers and a pooling layer, organized in four parallel paths of the information flow. It emerged as a component of CNN architecture called GoogLeNet, which has about 100 convolutional layers, distributed in nine Inception v3 modules, organized in 22 depth layers (Szegedy et al., 2015). Although its use in deep CNN is well established, little is known about the possible effects of the module on neural networks with a reduced number of convolutional layers. In this context, this study aimed to build a Computational Vision System to classify the degree of browning of bread during baking, based only on the color changes of the bread crust, using a Convolutional Neural Network as a base, composed of a reduced number of convolutional layers and the Inception v3 module (Short-CNN), maintaining low consumption of computational resources.

2. MATERIALS AND METHODS

2.1. Samples preparation

Bread samples were prepared according to the formulation described by Purlis and Salvadori (2009). The dough was prepared by mixing the ingredients for 15 minutes in a stainless steel mixer (G-Paniz, BPR-5L) at 120 rpm. Individual samples of 50 g were mechanically modeled in a bread cylinder (G-Paniz, CL390) and kept in a fermentation chamber, with temperature and moisture control, for approximately three hours for expansion.

Table 1. Codification of the baking stages of the bread samples based on the respective baking times

Time (min.)	Stages
0	Raw
5	Unbaked-1
10	Unbaked-2
15	Baked
20	Overbaked-1
25	Overbaked-2
30	Burnt

In order to avoid any possible effects of position in the oven on the browning of the bread crust, the samples were evenly distributed on the tray. The trays were positioned in the upper and lower half of the oven. The samples were baked in an electric oven (G-Paniz, FTE300) under forced convection at 180 °C (±5 ºC) for seven different baking periods (0, 5, 10, 15, 20, 25 and 30 min.). After each baking period, the samples were cooled for two minutes and then photographed in the acquisition chamber. The images obtained were coded, defining the seven baking stages (Table 1). For each baking stage, at least 17 samples were prepared in three repetitions. A total of 374 samples were prepared.

2.2. Acquisition of images

Image acquisition was performed using a digital camera (OLYMPUS - VG120, D705) positioned 25 cm from the sample. Each photo was taken with a 2.6x fixed optical zoom. A 25 x 25 x 20 cm (width x height x depth) white background illumination chamber was used equipped with an 8 W (6500 K) fluorescent lamp (Figure 1). Images were captured with color representation in sRGB space (Standard RGB), 314 dpi resolution and dimensions of 4,288 x 3,216 pixels.

2.3. Preprocessing of images

To ensure that the neural network would extract characteristics based on the color and texture patterns of the bread crust, an image fragment of 600 x 600 pixels in the crust region was manually extracted from each image (Figure 1).
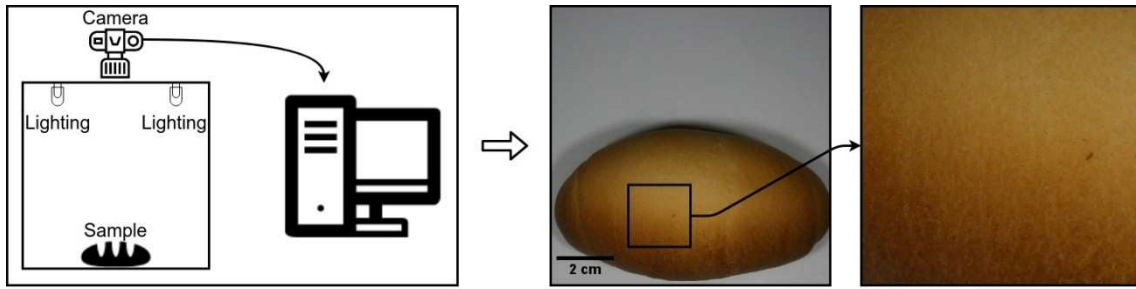
Figure 1. Scheme of the image capture system and the image cutting process in the bread crust region.

The region of interest was carefully chosen so that the elements of bread shape, such as edges and cracks in the crust, as well as the background of the image were completely eliminated (Figure 2). The entire image editing operation was performed in the GNU Image Manipulation Program (GIMP 2.10.8).

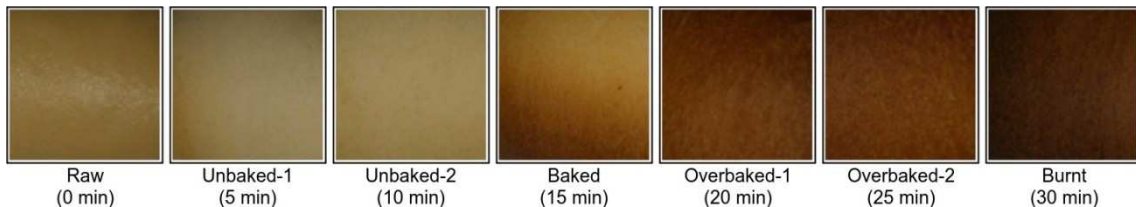| Raw (0 min) | Unbaked-1 (5 min) | Unbaked-2 (10 min) | Baked (15 min) | Overbaked-1 (20 min) | Overbaked-2 (25 min) | Burnt (30 min) |

Figure 2. Example of images of each of the seven baking stages (times) that make up the dataset used for training, validation, and testing of CNN.

2.4. Architecture of Convolutional Neural Networks

In this work, we proposed the use of a Short Convolutional Neural Network (Short-CNN), which was compared to the classical architectures (AlexNet and VGG-16) in terms of its accuracy and loss.

2.4.1. Short Convolutional Neural Network

Figure 3 shows a scheme of the Short-CNN. This was composed of two convolutional layers with a normalization step (Batch normalization), interleaved by pooling layers, which is responsible for downsizing the image (downsampling) (Muñoz et al., 2019). The first convolutional layer consists of 96 convolutional kernels (7x7) and the second by 144 kernels (3x3). Next, a four-way Inception v3 module, followed by a third convolutional layer, connects to a fully connected region. The graphical representation of the Inception v3 module is in the center of Figure 3. The Inception v3 module uses the same equations described as the other convolutional layers. All layers of the Inception module comprise 36 convolutional kernels. In a fully connected region, the first hidden layer has 256 neurons and the second 128 neurons. Both are followed by a regularization step (Dropout). Continuing, there is an output layer composed of seven neurons, according to the number of stages defined in Table 1. The Softmax function was used as an activation function because it is an image classification problem and is often used due to its characteristic of producing a distribution of probabilities of the output value belonging to a particular class (Oh and

Kim, 2017). All other layers used the activation function "Rectified Linear Units" (ReLU). The network training was implemented in single GPU (Graphics Processing Unit) processing. In order to maintain consistency with the other architectures employed, the images used in training, validation, and testing were resized to a size of 227 x 227 x 3 pixels.
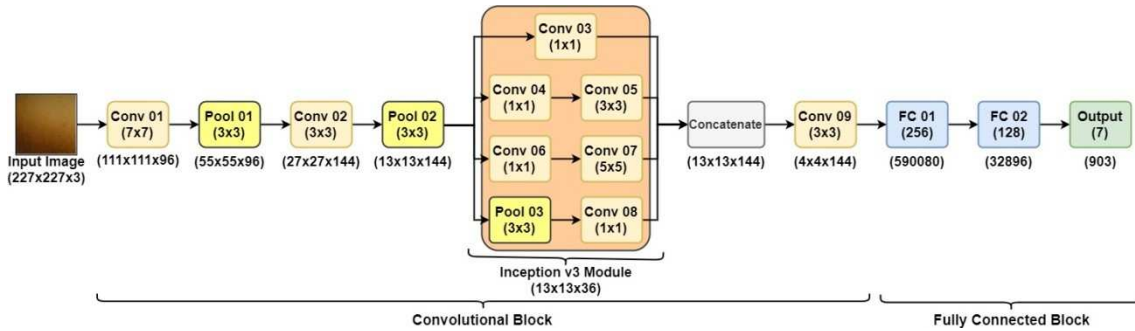

Figure 3. Graphic representation of the Short-CNN containing an input data layer, a Convolutional Block with embedded Inception v3 module, and a Fully Connected Block containing the output layer.

2.4.2. AlexNet

The AlexNet architecture, which has already been used to solve problems involving colors (Oh and Kim, 2017), was adopted with some modifications. The implementation of dual GPU processing has been replaced by single GPU processing. The images used in training, validation and testing were resized to the size 227 x 227 x 3 pixels. The number of neurons in the output layer was equal to the number of baking stages. The pooling operation, after the first two convolutional layers, was performed by applying the "MaxPooling" function. The normalization of the values obtained after each convolutional layer was performed by the "Batch Normalization" technique (Ioffe and Szegedy, 2015) replacing the "Local Response Normalization" technique since it has proven to be more advantageous than this from the viewpoint of memory consumption and computational time (Bayar and Stamm, 2017; Simonyan and Zisserman, 2015). In order to reduce the overfitting effect, the Dropout regularization technique was applied in layers one and two of the fully connected section of the neural network. In all hidden layers, as well as in the first layer, the activation function "Rectified Linear Units" (ReLU) was adopted. In the output layer, the Softmax function was adopted.

2.4.3. VGGNet-16

The VGGNet-16 architecture (Simonyan and Zisserman, 2015), was composed of 16 layers, distributed in six blocks. Similar to what was observed for the AlexNet architecture, the implementation of multiple GPU processing was replaced by single GPU processing. The images used in training, validation and testing were resized to the size 227 x 227 x 3 pixels. The number of neurons in the output layer was equal to the number of baking stages. The first block, composed of two convolutional layers with 64 convolutional kernels (3x3) each, was followed by a

pooling layer (MaxPooling). The second block was composed of two convolutional layers with 128 kernels (3x3) also followed by a pooling layer (MaxPooling). The third block was composed of three convolutional layers with 256 kernels (3x3) each, followed by a pooling layer (MaxPooling). The fourth and fifth blocks presented three convolutional layers with 512 kernels (3x3), followed by a pooling layer (MaxPooling). The sixth block was composed of four fully connected layers. The first two were composed of 4096 neurons. The second to last layer presents 1000 neurons. In all of them, the ReLU activation function was used, except in the last layer which was used the softmax function.

2.5. Convolutional neural network training strategy

The initial dataset was composed of 374 images distributed in seven stages of baking. The dataset was randomly divided into three subsets: training (70%), validation (15%) and test (15%). In order to reduce the risk of overfitting and increase the generalization capacity of CNN, "Data Augmentation" technique was applied to the dataset (Gu et al., 2018). The data augmentation was performed through the combined use of geometric transformation methods of translation, zoom, mirroring and image rotation (-45° to +45°, variation of 1°). The size of the dataset was increased by the random combination of geometric transformations by an average factor of 19, a value similar to that used by Chen et al. (2019) and lower than the used by Lathuilière et al. (2019). After applying the "Data Augmentation" technique, the dataset reached a total of 7158 images, distributed among the baking stages (Table 2).

The network hyperparameters were updated by the Adam Optimizer algorithm (Kingma and Ba, 2015) with a learning rate of 1×10^{-3}, decay of the learning rate of 1×10^{-3} and batch size equal to 12 images. The initialization of convolutional kernels matrices was performed by the technique known as Glorot Uniform (Glorot and Bengio, 2010). In order to ensure that the CNN performance is not influenced by the sequence of appearance of the images in the dataset, the technique of randomization (shuffling) of training dataset was applied before each training epochs (Ioffe and Szegedy, 2015; Meng et al., 2019). We randomized the training data 200 times during the entire process.

Table 2. Distribution of the dataset among the subsets of training, validation, and testing, as well as among the seven stages corresponding to the seven baking periods

Stage	Training	Validation	Test
Raw	736	150	150
Unbaked-1	704	154	154
Unbaked-2	714	152	152
Baked	720	153	153
Overbaked-1	720	153	153
Overbaked-2	703	152	152
Burnt	725	154	154
Total	**5022**	**1068**	**1068**

The matrices with the weights (W) of the convolutional neural network were adjusted by minimizing the loss function "Categorical Cross-Entropy Loss". For a relationship of N pairs of desired input-output data $\{(\boldsymbol{x}^{(n)}, \boldsymbol{d}^{(n)}); n \in [1, 2, ..., N]\}$, where $\boldsymbol{x}^{(n)}$ is the nth input data and $\boldsymbol{d}^{(n)}$ the corresponding nth output data desired, with $\boldsymbol{y}^{(n)}$ representing the nth output data of the CNN, $\boldsymbol{\theta}$ a given parameter to be adjusted, the cost function to be minimized was the loss function ($\mathcal{L}$), which can be generalized as (Equation 1):

$$\mathcal{L} = \frac{1}{N} \sum_{n=1}^{N} \ell(\boldsymbol{\theta};\, \boldsymbol{d}^{(n)}, \boldsymbol{y}^{(n)}) \qquad (1)$$

Training, validation and testing procedure of the neural network were implemented through a script written in Python 3.6 language, with the TensorFlow 1.12, Keras 2.2, OpenCV4.0, Matplotlib 2.2 and Numpy 1.16 libraries, using a microcomputer equipped with Intel Core i5 6500 processor, with 24 GB of memory and Nvidia GeForce RTX 2060 video card.

2.6. Metrics for performance evaluation of CNN

The performance of the Short-CNN, AlexNet and VGGNet-16 networks was evaluated by monitoring the minimization of the Loss Function and the evolution of Global Accuracy. In addition, for the Short-CNN, the Accuracy (Equation 2), Precision (Equation 3), Recall (Equation 4) and F1 Score (Equation 5) of each of the stages corresponding to the different baking times were also evaluated (Markovic et al., 2018).

$$Accuracy = \left(\frac{TP + TN}{TP + TN + FP + FN}\right) \qquad (2)$$

$$Precision = \left(\frac{TP}{TP + FP}\right) \qquad (3)$$

$$Recall = \left(\frac{TP}{TP + FN}\right) \qquad (4)$$

$$F1 = \left(\frac{2TP}{2TP + FP + FN}\right) \qquad (5)$$

where TP (True Positive) is the number of images correctly classified as positive at a particular baking stage, TN (True Negative) the number of images correctly classified as negative at the same baking stage, FP (False Positive) the number of images incorrectly classified as positive at this baking stage, and FN (False Negative) the number of images incorrectly classified as negative at that baking stage.

3. RESULTS AND DISCUSSION

In the construction of the CVS, we sought to combine the ability to discriminate the color of the bread crust during the baking process with low computational resource consumption. In this sense, the Short-CNN presented low memory size requirement allocated to the training stage, representing a reduction of 91.6% compared to the AlexNet architecture and 97.0% in relation to VGGNet-16 (Table 3).

Table 3. Comparison of the Short Convolutional Neural Net (Short-CNN) structure and the AlexNet and VGGNet-16 architectures

Layers	Short-CNN	AlexNet	VGGNet-16
Input	227 x 227 x 3	227 x 227 x 3	227 x 227 x 3
Layer 01	CL, 96	CL, 96	CL, 64
Layer 02	CL, 144	CL, 256	CL, 64
Layer 03	*IC v3*, 36 (144)	CL, 384	CL, 128
Layer 04	CL, 144	CL, 384	CL, 128
Layer 05	-	CL, 256	CL, 256
Layer 06	-	-	CL, 256
Layer 07	-	-	CL, 256
Layer 08	-	-	CL, 512
Layer 09	-	-	CL, 512
Layer 10	-	-	CL, 512
Layer 11	-	-	CL, 512
Layer 12	-	-	CL, 512
Layer 13	-	-	CL, 512
Layer 14	FC, 256	FC, 4096	FC, 4096
Layer 15	FC, 128	FC, 1024	FC, 4096
Output	FC, 7	FC, 7	FC, 7
Memory*	**3.9 M**	**46.3 M**	**128.3 M**

CL: convolutional layer; FC: fully connected; IC: Inception Module
*Estimated memory.

Three basic premises were adopted in the elaboration of the Short-CNN. First, reduction in the number of convolutional layers, which implies a reduction in the number of active layers, reducing the number of nonlinear operations, reducing the complexity of the model and enabling the increase in the network training speed (Q. Zhang et al., 2018). Second, reduction in the number of convolutional kernels, which leads to a reduction in the number of adjustable parameters of the network, enabling the training speed increase and reducing the risk of overfitting (Wu, 2018). Finally, the use of the Inception v3 module enables a higher number of convolutional layers to be maintained, increasing the width of the network instead of the depth. Additionally, the presence of 1x1 kernels enables the control of the growth of the number of adjustable network parameters. Although Short-CNN presents a total of nine convolutional layers, a value higher than that observed in AlexNet architecture, the amount of adjustable parameters is 12 times lower.

Figure 4 shows the performance of Short-CNN compared to the two classic architectures (AlexNet and VGGNet-16), measured by the evolution of global accuracy and loss function during the training and validation process. As observed, the reduction in the number of stacked convolutional layers, as well as the reduction in the number of convolutional kernels and the use of the Inception v3 module enabled the increase in the network training speed. After 30 training epochs, the Short-CNN reached global accuracy values of over 90%, both with the training and validation dataset. The AlexNet architecture only reached such values after 100 training epochs and the VGGNet-16 architecture was below this value in the validation even after 200 training epochs.

In order to verify the generalization capacity of the convolutional neural network, an additional test dataset was used. In this step, after 200 training epochs the Short-CNN presented global accuracy of 98.8%, a value higher than that presented by AlexNet (98.1%) and VGGNet-16 (92.5%). It is important to emphasize that the reduction in model size did not affect its ability to discriminate the different stages of bread baking. This is particularly important because it can allow the use of such network architecture in equipment with low processing capacity and data storage, as is the case of mobile devices or embedded systems (Tran et al., 2018).
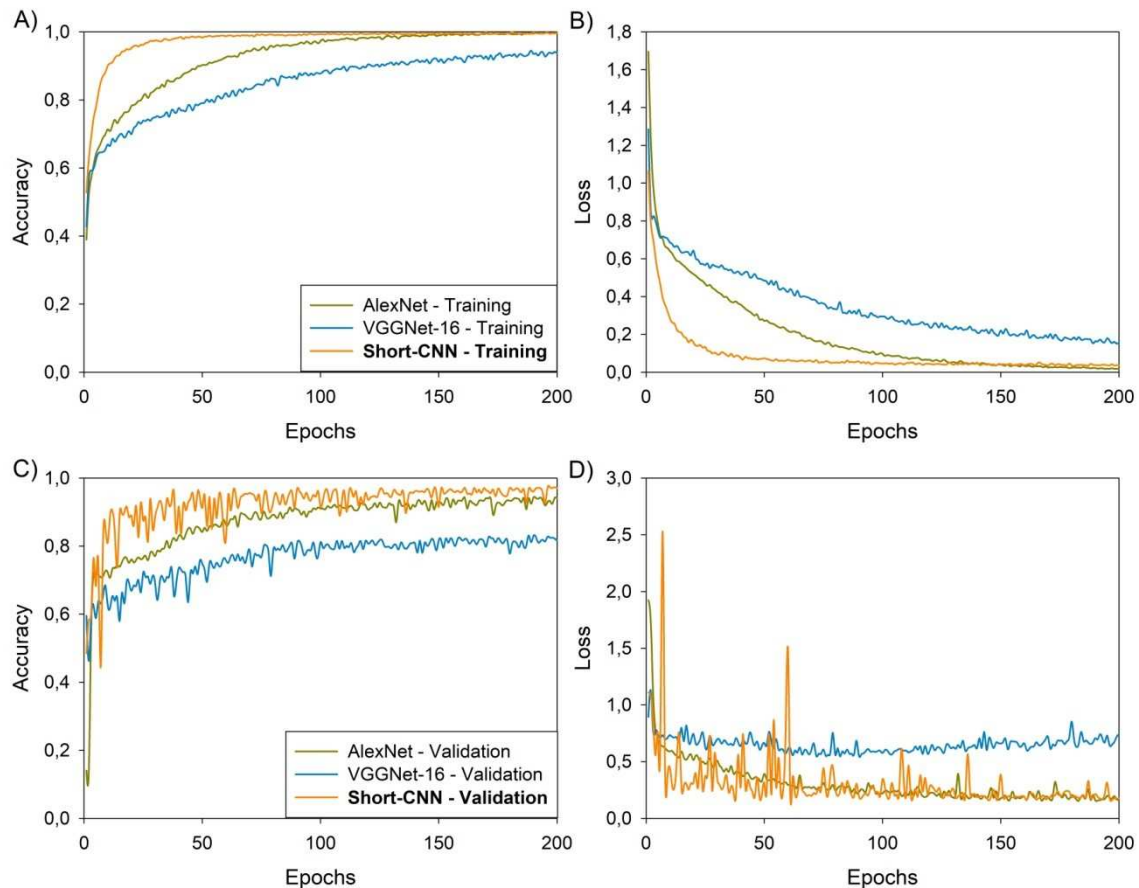

Figure 4. Comparative performance of Short-CNN versus two classical architectures, AlexNet and VGGNet-16, measured by the evolution of A) Training Accuracy, B) Training Loss, C) Validation Accuracy and D) Validation Loss over epochs.

During the baking process, the crust of the bread undergoes a browning process, caused by the production of colored substances due to the Maillard and caramelization reactions. The rate of browning is slow in the initial stage of baking and intensifies as the processing time evolves (Helou et al., 2016). Therefore, a good CVS should be able to distinguish the different baking stages, both in the initial phase when the color changes are still slow, and also in the final stage, when great changes in the color of the bread surface ensue. A Confusion Matrix (Figure 5) shows the ability of the proposed model to correctly discriminate each of the stages.

In the test dataset, the Short-CNN was able to correctly identify 100% of the images of stages one, four and six (Raw, Baked, and Overbaked-2). Stages two and three (Underbaked-1 and Underbaked-2) presented a precision of 98.1% and 99.3%, respectively. Stages five and seven (Overbaked-1 and Burnt) presented a hit rate higher than 96.5%. In ordinary CVS, the deviations were not uniformly distributed throughout the processing time but are concentrated in certain regions, which is one of the main limitations of those models (Broyart and Trystram, 2003; Putranto et al., 2015). In the Short-CNN the deviations are uniformly distributed over the baking stages, showing its ability to discriminate the stages of baking bread with a high degree of accuracy, regardless of the processing time.
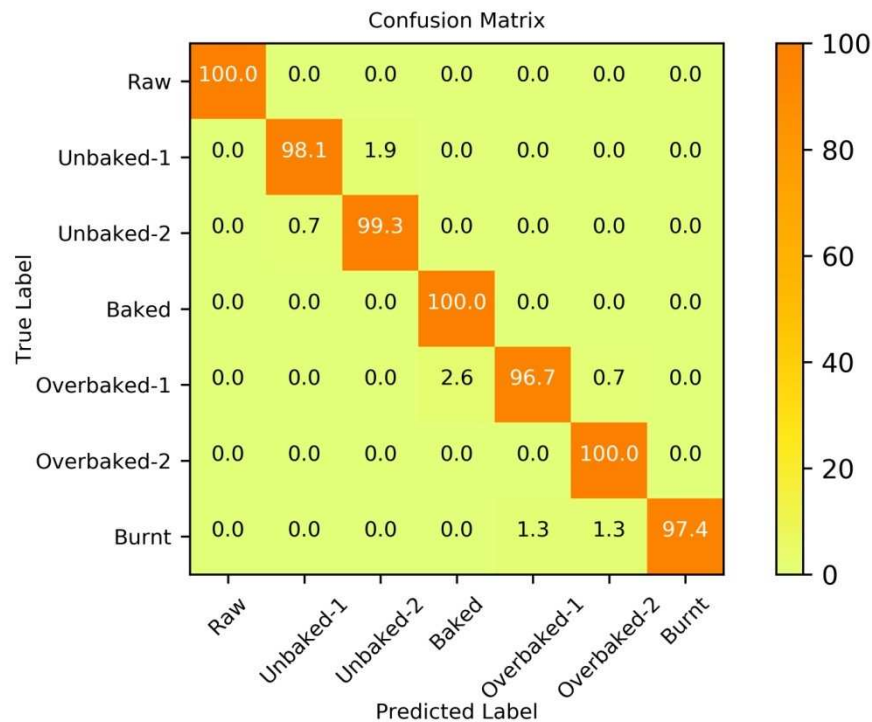

Figure 5. Confusion matrix for Short Convolutional Neural Net (Short-CNN) diagonally showing the fraction of correct predictions for each bread baking stage.

Table 4 shows individualized indicators by class. The use of indicators segmented by class aims to demonstrate that Short-CNN does not privilege one class over the others (Goutte and Gaussier, 2005). As can be seen in the accuracy score, the Short-CNN presents a frequency of correct prediction higher than 99.0%

for all classes. When evaluating the precision score, that is, of those predictions classified as correct, how many were effectively correct, all classes presented values higher than 97.4%.

The recall indicator measures the frequency of classification of an image in a specific class since it effectively belongs to that class. The Short-CNN correctly classified all images of Raw, Baked and Overbaked-2 classes. The Overbaked-1 class was that with the lowest value for this indicator (96.7%). The balance between precision and recall shows the model's ability to produce reliable results, being measured by the F1 Score indicator. As observed, for all classes the value of this indicator was over 97.5%, confirming the ability of the Short-CNN to correctly classify each of the images of the test group, regardless of the class to which they belong.

In order to verify the ability of the Short-CNN to extract color features from the images, the kernels of the first convolutional layer with homogeneous colors were selected by visual inspection. It was found that 22.9% of the convolutional kernels showed some degree of color selectivity. This value is lower than that observed in the literature (Rafegas and Vanrell, 2018), where 39.6% of the kernels of the first convolutional layer showed some degree of selectivity for colors. However, different from the study of those authors, in the present paper, the images are composed only of the bread crust region, eliminating other texture elements present in them, which hinders the training of the neural network. Figure 6 shows the kernels of the first convolutional layer selective for colors.

Table 4. Accuracy, precision, recall and F1 Score, segmented by class, for the Short-CNN using the test dataset

Stages	Accuracy	Precision	Recall	F1 Score
Raw	1,0000	1,0000	1,0000	1,0000
Unbaked-1	0,9962	0,9934	0,9805	0,9869
Unbaked-2	0,9962	0,9805	0,9934	0,9869
Baked	0,9962	0,9745	1,0000	0,9871
Overbaked-1	0,9934	0,9867	0,9673	0,9769
Overbaked-2	0,9972	0,9806	1,0000	0,9902
Burnt	0,9962	1,0000	0,9740	0,9868

The Short-CNN presented selective kernels for the crust colors of the bread samples present in each of the seven baking stages. Similar results were also observed in other works (Q. Zhang et al., 2018; Rafegas and Vanrell, 2018). These authors also observed that there is a strong association between the extraction of color features and texture features. Such characteristic is evident in Figure 6, where it is observed that each convolutional kernel extracted simultaneously color and texture features, representing the non-homogeneity of the color distribution present on the bread crust (Figure 1). This information shows the ability of the CNN to classify images based on a color feature map in a similar manner as the texture feature map (Gu et al., 2018). This behavior of CNN is strongly correlated with the human vision system (Rafegas and Vanrell, 2018). Thus, the color feature map

extracted by Short-CNN can be used to identify and classify baking stages of bread even if they have been obtained under different conditions. Therefore, it is enough that the cooking stage in question has the same texture and color pattern of a stage previously learned by Short-CNN, i.e., it is not necessary to know the temperature or the processing time used to reach that stage. The ability of the model to extract colors present in the samples, from white to dark brown, combined with the fact that it does not depend on temperature or process time, allows it to be used in other conditions, such as different sample sizes, different types and oven temperatures and even, with some limitations, in different formulations.
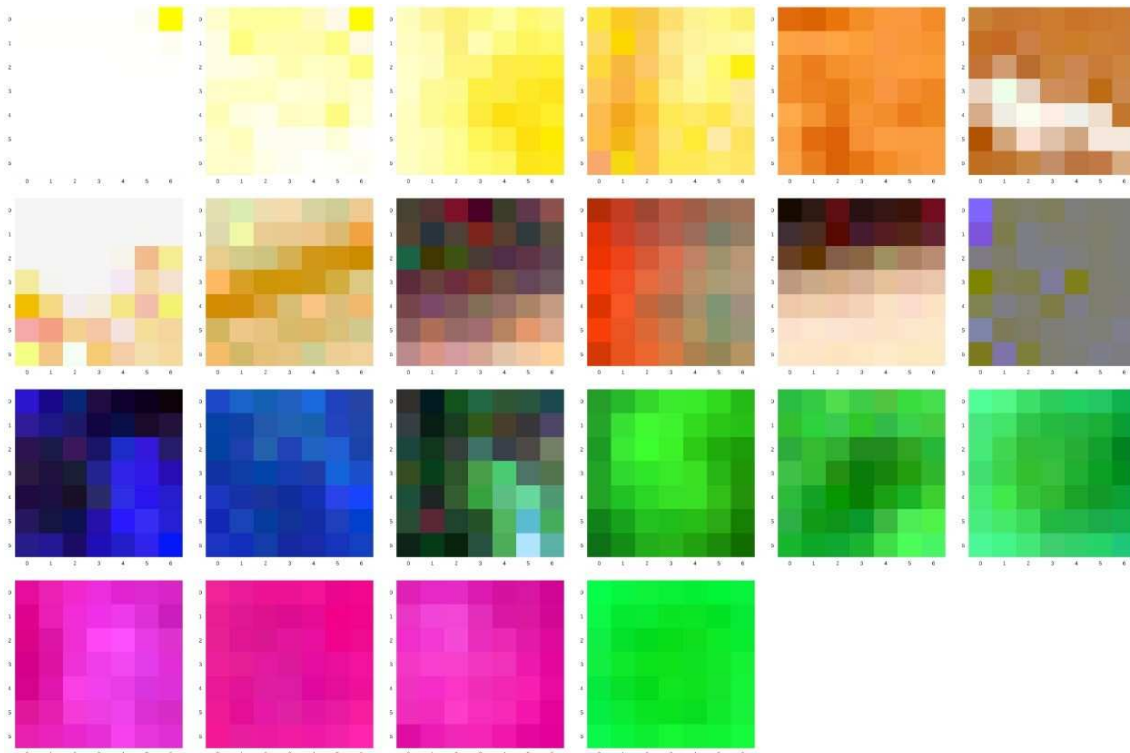

Figure 6. Color-selective kernels of the first convolutional layer of Short-CNN.

In addition, there is the presence of selective convolutional kernels for other colors (blue, pink and green) not perceptible by visual inspection in the samples. Such kernels were also observed by other authors (Rafegas and Vanrell, 2018), pointing to the possibility of using the resulting CVS to solve other problems involving colors, such as product acceptance prediction, classification of the degree of vegetable ripeness, enzymatic or non-enzymatic degree browning in food, degradation of pigments, etc.

4. CONCLUSIONS

In this paper, we presented a Computational Vision System (CVS), based on Convolutional Neural Networks (CNN), as a non-destructive tool for classification of the degree of browning in samples of bread during baking. The proposed CVS combines the ability to discriminate the degree of browning with non-invasive evaluation, eliminating the dependence on additional variables, as observed in

traditional CVS. In the implementation of the CVS, a Convolutional Neural Network was adopted with a reduced number of convolutional layers containing an Inception v3 module (Short-RNC) and using fragments of images of the bread crust during baking as input data.

The results obtained support the following conclusions. Firstly, the use of CNN allows the construction of a CVS not dependent on process variables, such as surface or oven temperature, water activity or sample moisture, which is a great advantage in relation to traditional models. Secondly, the use of a CNN with a reduced number of convolutional layers, and containing the Inception v3 module (Short-CNN), proved to be more advantageous than the use of classical architectures. The use of Short-CNN enables the reduction in the consumption of computational resources during the training stage allied to the increase in the discrimination capacity of the baking stages. Thirdly, the developed CVS presented homogeneous behavior, not being influenced by the color stage of the sample, overcoming a common limitation to the classical systems. Finally, the use of the Short-CNN did not compromise the recognition of the colors present in the samples, since in the first convolutional layer kernels selective to the colors present in the seven stages of baking were found. Additionally, other colors not identified in the samples were selected in convolutional kernels, which indicate the possibility of using the CVS to solve other problems involving colors, such as product acceptance prediction, classification of vegetable maturation degree, degree of enzymatic or non-enzymatic browning in food, pigment degradation, among others.

REFERENCES

Aarathi, K.S., Abraham, A., 2017. Vehicle color recognition using deep learning for hazy images. Proc. Int. Conf. Inven. Commun. Comput. Technol. ICICCT 2017 335–339. https://doi.org/10.1109/ICICCT.2017.7975215

Azman, A.A., Ismail, F.S., 2017. Convolutional Neural Network for Optimal Pineapple Harvesting. Elektr. J. Electr. Eng. 16, 1–4. https://doi.org/10.11113/eletrika.v16n2.54

Bayar, B., Stamm, M., 2017. Design Principles of Convolutional Neural Networks for Multimedia Forensics. Electron. Imaging 2017, 77–86. https://doi.org/10.2352/issn.2470-1173.2017.7.mwsf-328

Broyart, B., Trystram, G., 2003. Modelling of Heat and Mass Transfer Phenomena and Quality Changes During Continuous Biscuit Baking Using Both Deductive and Inductive (Neural Network) Modelling Principles. Food Bioprod. Process. 81, 316–326. https://doi.org/10.1205/096030803322756402

Castro, W., Oblitas, J., Chuquizuta, T., Avila-George, H., 2017. Application of image analysis to optimization of the bread-making process based on the acceptability of the crust color. J. Cereal Sci. 74, 194–199. https://doi.org/10.1016/j.jcs.2017.02.002

Chen, P.Y., Blutinger, J.D., Meijers, Y., Zheng, C., Grinspun, E., Lipson, H., 2019.

Visual modeling of laser-induced dough browning. J. Food Eng. 243, 9–21. https://doi.org/10.1016/j.jfoodeng.2018.08.022

Fogliano, V., Morales, F.J., 2011. Estimation of dietary intake of melanoidins from coffee and bread. Food Funct. 2, 117–123. https://doi.org/10.1039/c0fo00156b

Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. Proc. 13th Int. Conf. Artif. Intell. Stat. 9, 249–256.

Goutte, C., Gaussier, E., 2005. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. Lect. Notes Comput. Sci. 3408, 345–359. https://doi.org/10.1007/978-3-540-31865-1_25

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., Chen, T., 2018. Recent advances in convolutional neural networks. Pattern Recognit. 77, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

Guo, Y., He, Y., Song, H., He, W., Yuan, K., 2018. Correlational examples for convolutional neural networks to detect small impurities. Neurocomputing 295, 127–141. https://doi.org/10.1016/j.neucom.2018.03.017

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem, 770–778. https://doi.org/10.1109/CVPR.2016.90

Helou, C., Jacolot, P., Niquet-Léridon, C., Gadonna-Widehem, P., Tessier, F.J., 2016. Maillard reaction products in bread: A novel semi-quantitative method for evaluating melanoidins in bread. Food Chem. 190, 904–911. https://doi.org/10.1016/j.foodchem.2015.06.032

Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, in: Bach, F., Blei, D. (Eds.), Proceedings of the 32 Nd International Conference on Machine Learning. Lille, pp. 448–456.

Kingma, D.P., Ba, J., 2015. Adam: A Method for Stochastic Optimization, in: International Conference on Learning Representations (ICLR 2015). Ithaca, NY: arXiv.org, San Diego, pp. 1–15.

Kocadağli, T., Gökmen, V., 2016. Multiresponse kinetic modelling of Maillard reaction and caramelisation in a heated glucose/wheat flour system. Food Chem. 211, 892–902. https://doi.org/10.1016/j.foodchem.2016.05.150

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks, in: Advances in Neural Information Processing Systems 25 (NIPS 2012). pp. 1–9. https://doi.org/10.1016/B978-008046518-0.00119-7

Lathuilière, S., Mesejo, P., Alameda-pineda, X., Horaud, R., 2019. A Comprehensive Analysis of Deep Regression. IEEE Trans. Pattern Anal. Mach. Intell. 41, 1–17. https://doi.org/10.1109/TPAMI.2019.2910523

Li, H.X., Si, H., 2017. Control for Intelligent Manufacturing: A Multiscale Challenge. Engineering 3, 608–615. https://doi.org/10.1016/J.ENG.2017.05.016

Markovic, R., Grintal, E., Wölki, D., Frisch, J., van Treeck, C., 2018. Window opening model using deep learning methods. Build. Environ. 145, 319–329. https://doi.org/10.1016/j.buildenv.2018.09.024

Meng, Q., Chen, W., Wang, Y., Ma, Z.-M., Liu, T.-Y., 2019. Convergence analysis of distributed stochastic gradient descent with shuffling. Neurocomputing 337, 46–57. https://doi.org/10.1016/j.neucom.2019.01.037

Muñoz, I., Gou, P., Fulladosa, E., 2019. Computer image analysis for intramuscular fat segmentation in dry-cured ham slices using convolutional neural networks. Food Control 106, 106693. https://doi.org/10.1016/j.foodcont.2019.06.019

Oh, S.W., Kim, S.J., 2017. Approaching the computational color constancy as a classification problem through deep learning. Pattern Recognit. 61, 405–416. https://doi.org/10.1016/j.patcog.2016.08.013

Purlis, E., 2012. Baking process design based on modelling and simulation: Towards optimization of bread baking. Food Control 27, 45–52. https://doi.org/10.1016/j.foodcont.2012.02.034

Purlis, E., Salvadori, V.O., 2009. Modelling the browning of bread during baking. Food Res. Int. 42, 865–870. https://doi.org/10.1016/j.foodres.2009.03.007

Putranto, A., Chen, X.D., Zhou, W., 2015. Bread baking and its color kinetics modeled by the spatial reaction engineering approach (S-REA). Food Res. Int. 71, 58–67. https://doi.org/10.1016/j.foodres.2015.01.029

Rafegas, I., Vanrell, M., 2018. Color encoding in biologically-inspired convolutional neural networks. Vision Res. 151, 7–17. https://doi.org/10.1016/j.visres.2018.03.010

Schmittmann, O., Lammers, P., 2017. A true-color sensor and suitable evaluation algorithm for plant recognition. Sensors (Switzerland) 17. https://doi.org/10.3390/s17081823

Shen, Y., Zhou, H., Li, J., Jian, F., Jayas, D.S., 2018. Detection of stored-grain insects using deep learning. Comput. Electron. Agric. 145, 319–325. https://doi.org/10.1016/j.compag.2017.11.039

Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition, in: Bengio, Y., LeCun, Y. (Eds.), 3rd International Conference on Learning Representations, ICLR 2015. San Diego, pp. 1–14.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 07-12-June, 1–9. https://doi.org/10.1109/CVPR.2015.7298594

Tran, D.T., Iosifidis, A., Gabbouj, M., 2018. Improving efficiency in convolutional neural networks with multilinear filters. Neural Networks 105, 328–339. https://doi.org/10.1016/j.neunet.2018.05.017

Vanin, F.M., Lucas, T., Trystram, G., 2009. Crust formation and its role during bread baking. Trends Food Sci. Technol. 20, 333–343.

https://doi.org/10.1016/j.tifs.2009.04.001

Varol, T., Canakci, A., Ozsahin, S., 2018. Prediction of effect of reinforcement content, flake size and flake time on the density and hardness of flake AA2024-SiC nanocomposites using neural networks. J. Alloys Compd. 739, 1005–1014. https://doi.org/10.1016/j.jallcom.2017.12.256

Wang, J., Ma, Y., Zhang, L., Gao, R.X., Wu, D., 2018. Deep learning for smart manufacturing: Methods and applications. J. Manuf. Syst. 48, 144–156. https://doi.org/10.1016/j.jmsy.2018.01.003

Wu, C.W., 2018. ProdSumNet: reducing model parameters in deep neural networks via product-of-sums matrix decompositions.

Wu, H., Zhao, J., 2018. Deep convolutional neural network model based chemical process fault diagnosis. Comput. Chem. Eng. 115, 185–197. https://doi.org/10.1016/j.compchemeng.2018.04.009

Zanoni, B., Peri, C., Bruno, D., 1995. Modelling of browning kinetics of bread crust during baking. LWT - Food Sci. Technol. 28, 604–609. https://doi.org/10.1016/0023-6438(95)90008-X

Zhang, L., Putranto, A., Zhou, W., Boom, R.M., Schutyser, M.A.I., Chen, X.D., 2016. Miniature bread baking as a timesaving research approach and mathematical modeling of browning kinetics. Food Bioprod. Process. 100, 401–411. https://doi.org/10.1016/j.fbp.2016.08.007

Zhang, Q., Zhuo, L., Li, J., Zhang, J., Zhang, H., Li, X., 2018. Vehicle color recognition using Multiple-Layer Feature Representations of lightweight convolutional neural network. Signal Processing 147, 146–153. https://doi.org/10.1016/j.sigpro.2018.01.021

Zhang, Y., Lian, J., Fan, M., Zheng, Y., 2018. Deep indicator for fine-grained classification of banana's ripening stages. Eurasip J. Image Video Process. 2018. https://doi.org/10.1186/s13640-018-0284-8

Zhong, R.Y., Xu, X., Klotz, E., Newman, S.T., 2017. Intelligent Manufacturing in the Context of Industry 4.0: A Review. Engineering 3, 616–630. https://doi.org/10.1016/J.ENG.2017.05.015

4. ARTIGO 3 - DEVELOPMENT OF A HYBRID SYSTEM BASED ON CONVOLUTIONAL NEURAL NETWORKS AND SUPPORT VECTOR MACHINES FOR RECOGNITION AND TRACKING COLOR CHANGES IN FOOD DURING THERMAL PROCESSING

Neste artigo será demonstrada a viabilidade de construção de sistemas híbridos compostos por redes neurais convolucionais e máquinas de vetores de suportes (HS CNN-SVM) para a classificação dos estágios de forneamento de pães. Será demonstrado que essa estratégia permite a redução no tempo de convergência ao mesmo tempo em que reduz a complexidade e, portanto, reduz o consumo de recursos computacionais, sem comprometer a capacidade do HS CNN-SVM em extrair características de cor das imagens. A seguir é reproduzido o inteiro teor do manuscrito publicado na *Chemical Engineering Science*.

Cotrim, W. da S., Felix, L. B., Minim, V. P. R., Campos, R. C., & Minim, L. A. (2021). Development of a hybrid system based on convolutional neural networks and support vector machines for recognition and tracking color changes in food during thermal processing. **Chemical Engineering Science**, 240, 116679. 385 https://doi.org/10.1016/j.ces.2021.116679

Development of a hybrid system based on convolutional neural networks and support vector machines for recognition and tracking color changes in food during thermal processing

Weskley da Silva Cotrim[1], Leonardo Bonato Felix[2], Valéria Paula Rodrigues Minim[3], Renata Cássia Campos[4], Luis Antônio Minim[5,*]

[1] Universidade Federal de Mato Grosso, *Campus* Universitário do Araguaia, Instituto de Ciências Exatas e da Terra, Avenida Valdon Varjão, nº 6390. Barra do Garças - Mato Grosso, Brasil, CEP: 78600-000, weskleycotrim@ufmt.br, https://orcid.org/0000-0002-8190-9632

[2] Universidade Federal de Viçosa, Departamento de Engenharia Elétrica, *Campus* Universitário – Viçosa – Minas Gerais, Brasil, CEP: 36570-000, leobonato@ufv.br, https://orcid.org/0000-0002-6184-2354

[3] Universidade Federal de Viçosa, Departamento de Tecnologia de Alimentos, *Campus* Universitário – Viçosa – Minas Gerais, Brasil, CEP: 36570-000, vprm@ufv.br, https://orcid.org/0000-0001-7143-2060

[4] Universidade Federal de Viçosa, Departamento de Engenharia Agrícola, *Campus* Universitário – Viçosa – Minas Gerais, Brasil, CEP: 36570-000, renata@ufv.br, https://orcid.org/0000-0002-5419-5064

[5] Universidade Federal de Viçosa, Departamento de Tecnologia de Alimentos, *Campus* Universitário – Viçosa – Minas Gerais, Brasil, CEP: 36570-000, lminim@ufv.br, https://orcid.org/0000-0002-1584-9117

*Corresponding author

Prof. Luis Antônio Minim

Universidade Federal de Viçosa, Department of Food Technology, *Campus* Universitário – Viçosa – Minas Gerais, Brasil, CEP: 36570-000, lminim@ufv.br

ABSTRACT

In order to build a non-destructive tool for the recognition and classification of bread baking stages, based exclusively on the color changes of the bread crust, the present work aims to propose the development of a Hybrid System (HS) composed of a Convolutional Neural Network (CNN) and a Support Vector Machine (SVM). For training, validation and testing of the HS 374 images of the bread crust were used, distributed over seven baking periods. The results showed that the HS CNN-SVM was able to correctly recognize and classify the baking stages without human intervention, overcoming even models based solely on CNN. In addition, the HS CNN-SVM reduced convergence time and memory consumption, favoring its use in mobile or embedded systems. Finally, the HS CNN-SVM maintained the ability to extract color map characteristics present in CNN, allowing its use in the construction of process control systems for the food industry in which color changes are involved.

Keywords: Deep Learning; Machine Learning; Color Change; Browning.

Abbreviations
ANN - Artificial Neural Networks

AUC - Area Under the Curve
CIELab - Commission Internationale de l'Éclairage Color Space
CNN - Convolutional Neural Network
CVS - Computational Vision System
GAP - Global Average Pooling
HS - Hybrid System CNN-SVM
MAC Operations - Multiplication and Accumulation Operations
RBF - Radial Basis Function
RGB – Red, Green, Blue Color Space
ReLU - Rectified Linear Unit
SV - Support Vectors
SVM - Support Vector Machine
sRGB - Standard RGB
ΔE - Color Difference

Conflict of interest
The authors have no conflicts of interest to declare that are relevant to the content of this article.

1. Introduction

Baking is a crucial step in the development of the sensory characteristics of many food products. During this stage, important sensory changes occur, such as changes in the texture, aroma, taste and color of the product. Modifications in color are particularly important as they are an indicator of the process stage and the quality of the final product. In baked foods, color changes occur due to non-enzymatic browning resulted mainly from Maillard's reaction (Bertrand et al., 2018; Helou et al., 2016). Due to their link with the development of aroma and flavor components, color changes have been used to build Computer Vision Systems (CVS) for quality monitoring and control in baked products (Broyart and Trystram, 2003; Putranto et al., 2015; Zhang et al., 2016).

The first Computer Vision Systems were based on first order polynomial models, in which the CIELab color space indexes (L*, a* and b*) were used to compose the color difference (ΔE) variable (Zanoni et al., 1995). Subsequently, the CVS were simplified and started to use only the luminosity component (L*), obtained by direct conversion of the three components of the RGB color space (Purlis and Salvadori, 2009). Such CVS presented good results for laboratory conditions. However, their practical use by the industry is limited, due to deviations observed for some temperature conditions, which can be attributed to nonlinearities characteristic of the process (Putranto et al., 2015). To overcome this limitation, models based on machine learning, such as the Artificial Neural Networks (ANN) were proposed (Broyart and Trystram, 2003). However, these models are dependent on variables

that are difficult to measure, such as surface temperature or humidity, or depend on external models to estimate such variables, as well as the high cost of sensors operating in the CIELab color space (Purlis and Salvadori, 2009; Schmittmann and Lammers, 2017; Zanoni et al., 1995). Add to this the challenges imposed by Industry 4.0, which requires the use of systems and intelligent equipment, in which information must be easily obtained and shared autonomously, allowing rapid decision making, with minimization of losses and costs (Li and Si, 2017; Wang et al., 2018; Zhong et al., 2017).

To overcome such limitations, approaches in the context of Deep Learning, such as Convolutional Neural Networks (CNN), seem to be the most promising way forward. CNN is known for its ability to recognize complex patterns present in images without the necessity of human intervention which meets industry 4.0 requirements (He et al., 2016; Krizhevsky et al., 2012; Simonyan and Zisserman, 2015). The main applications of CNN are associated with object detection and classification based on the recognition of shapes and textures in images (Guo et al., 2018; Shen et al., 2018; Wu and Zhao, 2018) or complex patterns present in multiple combined data sequences (Shi and Yang, 2019; Wu and Zhao, 2018; Yuan et al., 2020). However, recent studies have shown that CNN also can extract color characteristics in convolutional layers (Cotrim et al., 2020; Rafegas and Vanrell, 2018). This fact made their application possible in the development of recognition systems of the ripening stage in pineapple (Azman and Ismail, 2017) and bananas (Y. Zhang et al., 2018) and vehicle colors (Q. Zhang et al., 2018).

However, CNN demands a high computational cost, requiring the use of computer graphics units (GPU) for their implementation. In general, the higher the number of convolutional layers, the higher the computational cost (Gu et al., 2018; Wu and Zhao, 2018). In this scenario, there is a tendency to reduce both the depth and number of trainable parameters of CNN (Ide et al., 2020), which contributes to increase the training speed and reduce the risk of overfitting (Wu, 2018). Such premises allow the use of CNN in equipment with low processing and storage capacity, as is the case of embedded systems (Gao et al., 2018; Tran et al., 2018).

In this context, the feasibility of using CNN with reduced number of convolutional layers (Short-CNN) for recognition of non-enzymatic browning stages in the bread crust during baking has recently been demonstrated (Cotrim et al., 2020). Short-CNN showed a relative reduction of up to 97% in memory consumption and overall accuracy of 98.8%, overcoming classical architectures, despite the increased complexity of the model due to the insertion of the Inception v3 module. Thus, the convergence time (200 epochs) was negatively affected, since it is dependent on factors such as size, complexity and number of adjustable parameters of the model (Nagpal and Dubey, 2019). Although Short-CNN showed significant performance gains, the increase in convergence time points to a worsening in the model training performance. This result contradicts the current trend of optimizing the CNN training process, reducing its complexity and, consequently, reducing the classification time

of a single image (Ide et al., 2020; Oh et al., 2020). This strategy is vital to enable the use of CNN in mobile systems, embedded systems, and real-time processes (He and Sun, 2015; Ide et al., 2020). However, simply removing hidden layers from the classification region, or reducing the number of neurons in them, does not seem to be a viable option, as it compromises network performance (Nadian et al., 2015). Therefore, a possible solution to overcome this limitation is the adoption of a Hybrid System, in which CNN's ability to extract the map of characteristics from the image, without the necessity of human intervention, is combined with a good classification system, with low computational cost, as is the case of the Support Vector Machines (SVM). The application of SVM in the construction of classifiers is well known and documented, but always using human intervention for the selection of input data (Razzaghi et al., 2016; Tseng et al., 2016).

The SVM consists of a discriminatory classifier, formally characterized by an optimal hyperplane separating two linearly separable datasets (Battineni et al., 2019). The SVM classifier depends only on a small part of the dataset, the supporting vector data, for the positioning of the separation hyperplane. This makes it a powerful tool with low consumption of computational resources. However, complex datasets can lead to a sudden increase in training time and memory consumption (Nalepa and Kawulok, 2019), limiting the direct use of the color feature map extracted by CNN's convolutional layers. In this sense, the application of a function capable of reducing data complexity, such as Global Average Pooling (GAP), may present a possible solution to overcome such limitation (Hsiao et al., 2019). The GAP function reduces the tensor corresponding to the feature map of each convolutional layer to a vector, whose elements are given by the average value of spatial information present in the feature maps. This operation does not add any new adjustable parameters, and does not interfere with CNN's generalization capability (Lin et al., 2014).

The use of hybrid systems composed of CNN and SVM has been successfully applied to the recognition of 3D geometric shapes (Hoang et al., 2020), the classification of genetic data (Huynh et al., 2019) and the recognition of livestock brandings (Santos and Welfer, 2019). Evidence suggests that for small datasets, CNN-SVM Hybrid Systems outperform equivalent CNN models with Softmax classifier (Deepak and Ameer, 2020; Liu et al., 2018). In addition, it was also found that for limited datasets, the CNN-SVM HS proved advantageous over the Transfer Learning technique by exhibiting higher overall accuracy, reduced memory consumption, and reduction in the number of MAC (Multiplication and Accumulation operations) operations (Deepak and Ameer, 2020). Although the use of hybrid systems of CNN and SVM are well established for the recognition and classification of objects based on their shape and texture, little is known about their use for image recognition and classification based on color patterns. Thus, this study aimed to develop a Hybrid System CNN-SVM for classifying the degree of bread baking, based only on changes in color of the crust, with low computational resource consumption, which enables its application in mobile systems, embedded systems,

and real-time processes. CNN with low computational resource consumption enables the development of low-cost process control systems based on color changes for the food industry. Thus, this system can be applied in the development of automatic systems for controlling cooking processes, evaluating pigment degradation, predicting product acceptance, grading the degree of ripeness of vegetables, etc. The Hybrid System was composed by a Convolutional Neural Network (CNN), responsible for the extraction of the color feature map, a Global Average Pooling (GAP) function, responsible for the reduction of the size of the extracted feature map in each convolutional layer, and a Support Vector Machine (SVM), which is responsible for the classification of the samples, based on the dataset produced by the GAP function.

2. Material and Methods

2.1. Dataset

A dataset consisting of 374 images of the bread crust, obtained as described by Cotrim and collaborators (Cotrim et al., 2020), distributed in seven baking stages, was used. The seven baking stages correspond to seven different baking periods (0, 5, 10, 15, 20, 25 and 30 min) (Figure 1-C). Each baking stage has at least 51 images of the bread crust (600 x 600 pixels). The images were captured in a 25 x 25 x 20 cm (width x height x depth) white-backlit chamber equipped with an 8 W (6500 K) fluorescent lamp using a digital camera (OLYMPUS - VG120, D705) positioned at 25 cm from the sample (Figure 1-A). The images were captured with color representation in the sRGB space (Standard RGB) and resolution of 314 dpi.

To ensure that CNN extracts a feature map based solely on the bread crust color patterns, shape elements such as edges and cracks in the crust as well as the background of the image have been excluded (Figure 1-B). The entire image editing operation was performed in the GNU Image Manipulation Program (GIMP 2.10.8).
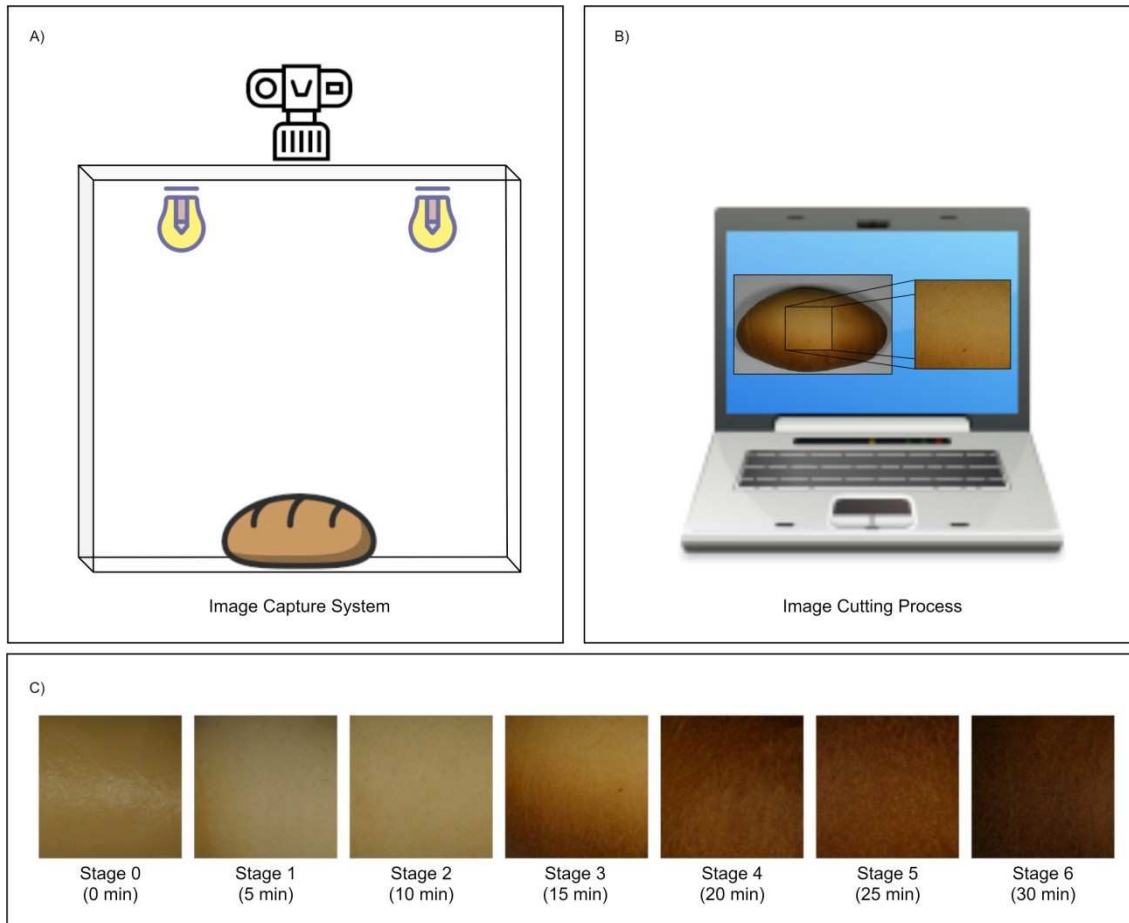
Figure 1. A) Image capture system, B) image cutting process in the bread crust region and C) example images of each of the seven baking stages (periods) that compose the dataset used for training, validation and testing of the Hybrid System (CNN-SVM).

2.2. Data augmentation

The initial dataset (374 images) was previously randomly divided into three subgroups: training (70%), validation (15%) and test (15%). In order to increase the generalization capacity and reduce the risk of overfitting, the Data Augmentation technique was applied to the dataset (Gu et al., 2018). Data augmentation was performed through the use of combined geometric transformation methods of translation, zoom, mirroring and image rotation (-45° to +45°, range 1°). The dataset was increased by the random combination of geometric transformations using a factor of 20. The choice of this factor was based on preliminary tests and is similar to those used by (Chen et al., 2019; Cotrim et al., 2020) and is smaller than that used by (Lathuilière et al., 2019). After applying the technique of data augmentation, the dataset increased to 7480 images.

2.3. Hybrid System Architecture (CNN-SVM)

The Hybrid System was structured to operate in two phases (Figure 2). In the first, composed by the Convolutional Neural Network (CNN), the extraction of the

feature map and its dimension reduction occurs by the application of the Global Average Pooling (GAP) function. In the second phase, composed by the Support Vector Machine (SVM), the image classification occurs. CNN was composed of three convolutional layers (7x7, 3x3 and 3x3), all with one normalization step (Batch Normalization). The first two convolutional layers were followed by two layers of Max Pooling (3x3). At the exit of each convolutional layer, the GAP function was applied, which were later reunited in a single layer by concatenation. The output of the concatenation layer was used as the input dataset for the SVM, organized in Bootstrap Aggregation with ten models. The result of the output layer of the Hybrid System was obtained by the composition of the result of the ten models in the "Voting" system.
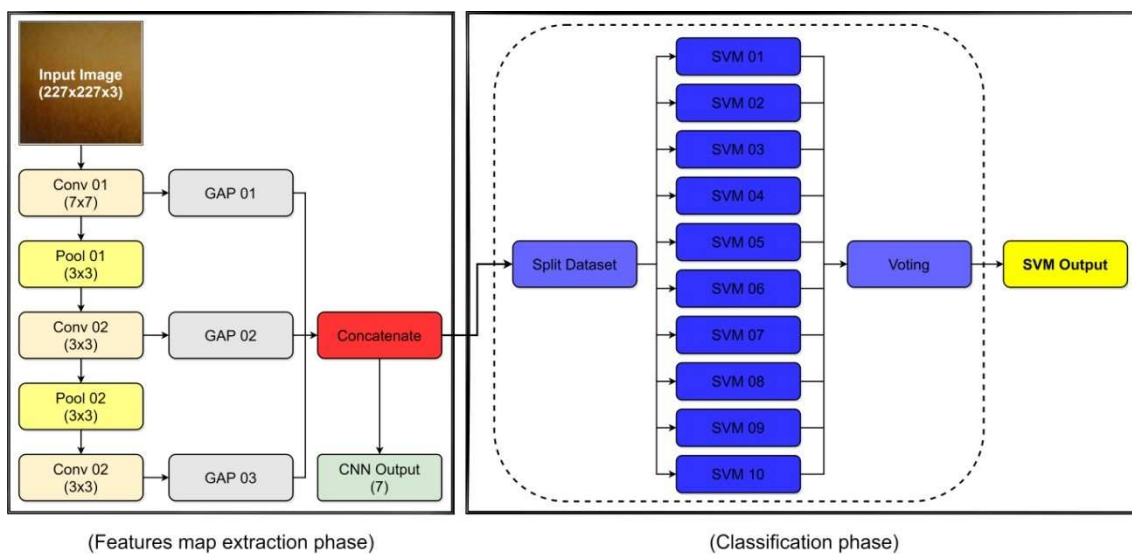

Figure 2. Graphical representation of the Hybrid System, composed by a Convolutional Neural Network in the feature map extraction phase and by a Support Vector Machine (Bootstrap Aggregation) in the classification phase.

2.4. Hybrid System Training Strategy

The dataset was randomly divided into two subgroups: training (85%) and test (15%). CNN hyperparameters were adjusted using the Adam Optimizer algorithm (Kingma and Ba, 2015), with learning rate of $1x10^{-3}$, learning rate decay of $1x10^{-3}$ and batch size equal to 12 images. The initialization of the convolutional kernels was performed using the Glorot Uniform technique (Glorot and Bengio, 2010). To ensure that the extraction of the feature map by CNN was not influenced by the order in which the images appeared in the dataset, the randomization technique (Shuffling) of the training dataset was used before each training epoch (Ioffe and Szegedy, 2015; Meng et al., 2019). The "Categorical Cross-Entropy Loss" function was adopted during CNN training as a loss function to be minimized.

After extraction of the feature map by CNN, the resulting concatenation layer from the GAP functions of each layer was used as input data for SVM training. The Bootstrap Aggregation strategy was adopted for this purpose, and the training

dataset was divided into 10 subgroups of equal size in a random manner, composed of 90% of the total training dataset. Each subgroup was used for training a SVM, totaling 10 SVM models. To train SVM the Radial Basis Function (RBF) was adopted with a penalty (C) equal to $1\text{x}10^3$ and the kernel coefficient (gamma) given by the "scale" function. The penalty (C) was adjusted according to the number of instances of each class present in the training dataset. To ensure that the model does not suffer interference from the dataset presentation order, the SVM training strategy with cross validation (Stratified k-fold) and k=10 was adopted.

The hybrid system training, validation and testing procedure was implemented using a script written in Python 3.6 language, with the libraries TensorFlow 1.12, Keras 2.2, OpenCV4.0, Matplotlib 2.2 and Numpy 1.16, using a microcomputer equipped with Intel Core i5 6500 processor, with 24 GB of memory and Nvidia GeForce RTX 2060 GPU.

2.5. Model evaluation metrics

Hybrid System performance was evaluated globally and in each class for accuracy, precision, revocation, F1 Score and Area Under the Curve (AUC) (Zhang et al., 2015). To ensure the consistency of the model, it was trained 10 times and the average value of each indicator obtained by the Micro-Average technique (Zhang and Zhou, 2014), as described by Equation 01.

$$\left.\begin{aligned}
Accuracy &= \frac{TP_j + TN_j}{TP_j + FP_j + TN_j + FN_j} \\
Precision &= \frac{TP_j}{TP_j + FP_j} \\
Recall &= \frac{TP_j}{TP_j + FN_j} \\
F1 &= \frac{2TP_j}{2TP_j + FP_j + FN_j} \\
AUC &= 1 - \int_0^1 f^{-1}\left(TP_{j,rate}\right)dTP_{j,rate}
\end{aligned}\right\} \quad (Eq.\,01)$$

where TP_j (True Positive) is the number of images of a class under evaluation correctly classified in the jth repetition, TN_j (True Negative) is the number of images of a class not belonging to the class under evaluation correctly classified in the jth repetition, FP_j (False Positive) is the number of images of another class incorrectly classified in the class under evaluation in the jth repetition and FN_j (False Negative) is the number of images of the class under evaluation incorrectly classified in another class in the jth repetition.

3. Theoretical basis

3.1. Theoretical basis for CNN

The extraction of the feature map in a CNN occurs by the operation between a convolutional kernel and the input data tensor of that layer. This convolutional kernel is trained by the backpropagation algorithm. Mathematically, the value of each z_{ijk}^l feature at position (i, j), at the kth position of the feature map, at the lth convolutional layer is given by (Equation 02):

$$z_{ijk}^l = {W_k^l}^T X_{ij}^l + b_k^l \qquad (Eq. 02)$$

where W_k^l is the weight tensor (convolutional kernel) of the kth convolution filter in the lth convolutional layer, b_k^l the bias of the kth convolution kernel in the lth convolutional layer and X_{ij}^l the fragment of the input tensor centered in position (i, j). Similar to what occurs in the artificial neural network (ANN), an activation function is applied to each element $\left(Z_{ijk}^l\right)$ of the tensor Z^l to introduce nonlinearities and detect nonlinear characteristics. The activation function mentioned here can be of the sigmoidal and hyperbolic tangent type (Gu et al., 2018) or, as is more common, of the ReLU (Rectified Linear Unit) type (Nair and Hinton, 2010). Mathematically, the general form of the activation function in the convolutional neural network can be given by (Equation 03):

$$a_{ijk}^l = a\left(z_{ijk}^l\right) \qquad (Eq. 03)$$

The pooling layer aims to obtain the shift invariance of the extracted feature by reducing the resolution (dimension) of the feature map tensor (Gu et al., 2018). There are many functions to perform the pooling such as "Average Pooling", "L_p Pooling", "Mixed Pooling", "Stochastic Pooling", among others. However, the most used is the "Max Pooling". In a generic way, the "pooling" operation can be represented as (Equation 04):

$$y_{ijk}^l = pool\left(a_{mnk}^l\right), \forall (m, n) \in \mathbb{R}_{ij} \qquad (Eq. 04)$$

where y_{ijk}^l is the kth element of the tensor of the lth pooling layer near the position (i, j). The stacking of successive layers of convolution and pooling is responsible for extracting high-level features from the dataset, allowing such information to be processed by the neural network in the "fully-connected" step, so that the desired solution is obtained. In classification problems it is common to use the Softmax function (Equation 05) as the activation function in the last layer of the neural network (Wu and Zhao, 2018).

$$\sigma\left(z_{ij}\right)_k = \frac{e^{z_i}}{\sum_{j=1}^k e^{z_j}}, j = 1, 2, \ldots, k \qquad (Eq. 05)$$

Convolutional neural network training is a global optimization problem, so it is necessary to establish a cost function to be minimized. For a relationship of N pairs of desired input-output data $\{(x^{(n)}, d^{(n)}); n \in [1, 2, \ldots, N]\}$, where $x^{(n)}$ is the nth input data and $d^{(n)}$ is the desired nth output data, with $y^{(n)}$ representing the nth output data of the CNN, θ a given parameter to be adjusted, the cost function to be minimized will be the loss function (L), which can be generalized as (Equation 06):

$$L = \frac{1}{N} \sum_{n=1}^{N} \ell\left(\theta; d^{(n)}, y^{(n)}\right) \qquad (Eq.\,06)$$

3.2. Theoretical basis for the SVM

In a SVM the classification occurs by the use of a decision hyperplane capable of distinguishing two linearly separable classes. For a linearly separable dataset, in $\mathbb{R}^{\mathcal{D}}$ space, the decision hyperplane is given by (Equation 07):

$$f(x): w^T \cdot x + b = 0 \qquad (Eq.\,07)$$

where w is the normal vector to the decision hyperplane, $w \in \mathbb{R}^{\mathcal{D}}$, and $b/\|w\|$ the perpendicular distance between the decision hyperplane and the origin, $b \in \mathbb{R}$ (Nalepa and Kawulok, 2019). In the case of two linearly separable classes, with classes labeled as $y_i \in \{-1, +1\}$, the training dataset must satisfy the condition (Equation 08):

$$y_i(w^T \cdot x_i + b) - 1 \geq 0, y_i \in \{-1, +1\} \qquad (Eq.\,08)$$

This equation represents two parallel vectors, positioned on the elements of each class closest to the decision hyperplane. The distance between each of these vectors and the decision hyperplane is given by $1/\|w\|$, so that the goal is to maximize the total distance between the two vectors, called separation margin. Thus, it is necessary that the vector w be minimized (Equation 09):

$$\min_{w,b} \frac{\|w\|^2}{2} \qquad (Eq.\,09)$$

Thus, the decision function corresponding to the support vectors (SV) can be expressed by (Equation 10):

$$f(a) = sgn\left(\sum_{i=1}^{t} \alpha_i y_i x_i^T a + b\right) \qquad (Eq.\,10)$$

where a is the vector of characteristics to be classified and α_i is the Lagrange multiplier. This form of decision function is highly sensitive to the presence of atypical data (outliers). To circumvent this limitation, a penalty (C) is applied, which admits some classification errors (ξ), but increases the generalization capacity of the model (Equation 11) (Cortes and Vapnik, 1995).

$$\min_{\boldsymbol{w},b,\xi} \frac{\|\boldsymbol{w}\|^2}{2} + C \sum_{i=1}^{t} \xi_i \qquad (Eq.\,11)$$

Finally, for non-linearly separable datasets, the kernel trick is applied, which consists of elevating the dimension of the data by applying a function ϕ, which allows obtaining a hyperplane capable of separating such classes (Equation 12).

$$\mathcal{K}(\boldsymbol{a}, \boldsymbol{a}') = \phi(\boldsymbol{a})^T \cdot \phi(\boldsymbol{a}') \qquad (Eq.\,12)$$

Thus, Equation 09 is rewritten into form (Equation 13):

$$f(\boldsymbol{a}) = sgn\left(\sum_{i=1}^{t} \alpha_i y_i \mathcal{K}(\boldsymbol{x}_i^T \cdot \boldsymbol{a}) + b \right) \qquad (Eq.\,13)$$

There are several types of kernel functions used in SVM for non-linearly separable data classification. Among them, the most used are the polynomial function and the Radial Basis Function (RBF) (Equation 14).

$$\mathcal{K}(\boldsymbol{a}, \boldsymbol{a}') = e^{\left(-\frac{\|\boldsymbol{a}-\boldsymbol{a}'\|^2}{2\sigma^2}\right)} \qquad (Eq.\,14)$$

3.3. Theoretical basis for the GAP

The Global Average Pooling (GAP) function is responsible for reducing the dimension of the feature map ($mxnxd$) of the lth layer, composed of the elements a_{ijk}^{l}, resulting in a tensor of dimension $1x1xd$, composed of the elements y_k^l (Lin et al., 2014). Mathematically, the value of each characteristic y_k^l in the kth position of the feature map, in the lth convolutional layer is given by (Equation 15):

$$y_k^l = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} \left[a\left(\boldsymbol{W}_k^{l\,T} \boldsymbol{X}_{ij}^l + b_k^l \right) \right] \qquad (Eq.\,15)$$

4. Results and Discussion

A new Hybrid System (HS) of computational vision, composed by a Convolutional Neural Network (CNN) and a Support Vector Machine (SVM), was proposed for recognition and classification of the bread baking stage, based solely on bread crust color changes, keeping the computational resources consumption low and with reduced convergence time. In the construction of HS, CNN had only 328,999 trainable parameters, representing a reduction of 67.5% compared to Short-CNN (1,012,304 trainable parameters) and 97.3% compared to AlexNet (12,185,148 trainable parameters) architecture (Cotrim et al., 2020). This reduction contributed to less memory consumption and reduced time required during the training stage, also reducing the risk of overfitting (Wu, 2018; Q. Zhang et al., 2018). In addition, HS CNN-SVM demanded less convergence time for CNN training, achieving performance compatible with the results of previous models, with only 20 training

epochs. This value represents a 90% reduction in convergence time compared to Short-CNN (200 epochs) during the training stage (Cotrim et al., 2020). The combination of the reduction in the number of adjustable parameters and the reduction in convergence time resulted in a reduction in total training time from 90 min to 6 min, and a reduction in computational processing time of 93.0%. The reduction in size, complexity, convergence and processing time minimizes the use of computational resources (Lin et al., 2019; Lym et al., 2019) and contributes to the reduction in energy consumption, which favors the use of CNN in mobile systems (Xu et al., 2020).

The Hybrid System was designed so that CNN extracts the color features map present in the crust of the samples for later use by SVM to classify the bread-baking stage. In order to evidence the ability of CNN to extract the color feature map, color selective convolutional kernels of the first convolutional layer were identified by visual inspection (Figure 3). At least 20.8% of the convolutional kernels of the first layer presented some degree of selectivity for the colors present in the bread crust after 20 epochs of CNN training. This result is similar to that observed by Cotrim and collaborators, where 22.9% of the convolutional kernels of the first convolutional layer presented some degree of selectivity for the colors present in the samples after 200 epochs of training (Cotrim et al., 2020).
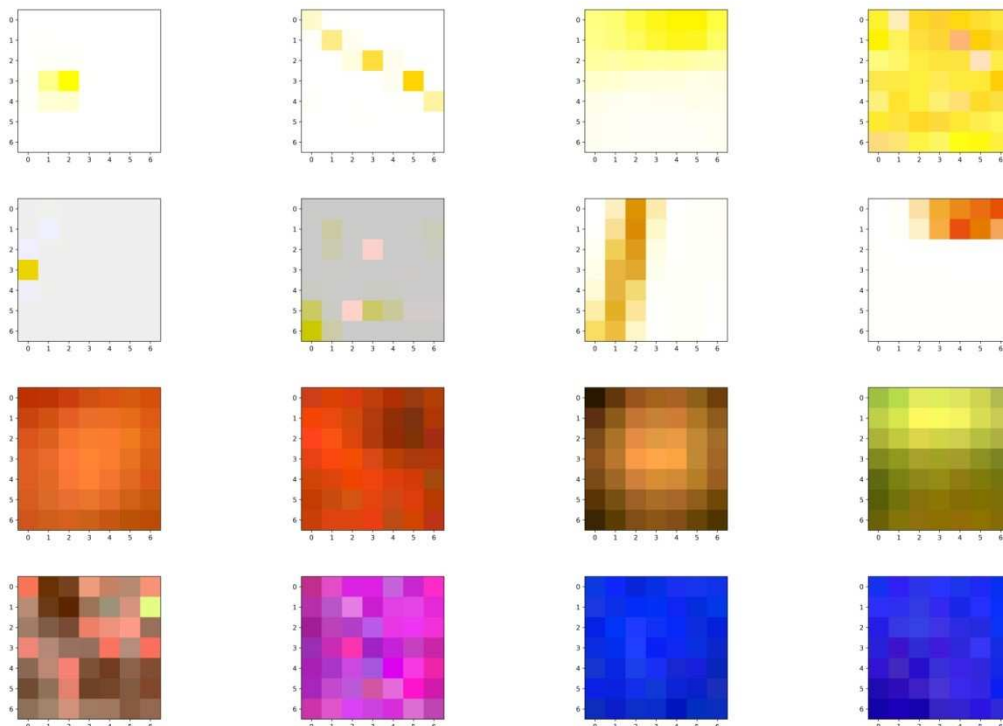

Figure 3. Color selective kernels of the first convolutional layer of the CNN-SVM Hybrid System.

These results show that CNN training with reduced convergence time (training epochs) does not affect its ability to extract color features map. This is particularly important as it ensures that the recognition and classification of baking stages by the

HS CNN-SVM will be based on the color characteristics present in the crust of the bread samples. Thus, texture elements such as edges and curves are discarded in this operation, ensuring that only the predominant colors in the samples contribute to the classification of the samples. This occurs due to the reduction in dimension of the feature map performed by the GAP function by calculating the mean value, converting the 2D tensor into a 1D tensor (Hsiao et al., 2019). Another important point to highlight is that the reduction in dimension of the tensors of the feature map is crucial to avoid that the SVM suffers an exponential increase in training time, as is common when the dataset is highly complex (Nalepa and Kawulok, 2019).

To ensure that the HS CNN-SVM was not influenced by the size of the dataset or the order of presentation of the images, the strategy of cross-validation was adopted during the training stage (Kohavi, 1995). The "Stratified k-Fold" algorithm with 10 subsets of data was adopted, which resulted in an overall average accuracy of 99.70% (±0.24%) of HS CNN-SVM during the training stage. The HS CNN-SVM generalization capacity was verified using the test dataset. There was a global average accuracy of 99.35% (±0.18%), with amplitude of 0.60%, minimum value 98.97% and maximum of 99.57% (Table 1). In addition, the HS CNN-SVM presented average global precision of 99.84% (±0.07%), average global recall of 99.35% (±0.18%), global F1 Score of 99.60% (±0.11%) and Area Under the Curve (AUC) 99.97% (±0.03%). To evaluate the contribution of CNN and SVM to HS CNN-SVM, a lateral output was introduced after the concatenation layer, using a Softmax classifier (Figure 3). In this case, the overall average accuracy of the CNN was only 84.20% (±4.20%), a value much lower than that obtained by the HS CNN-SVM (Figure 4), which shows that, in fact, the SVM contributes to the increase in the discriminatory power of the model. The combined use of CNN and SVM in the construction of a Hybrid System proved to be more efficient in recognizing and classifying the baking stages based only on the color present in the crust of the samples. This result indicates that HS CNN-SVM operating only with color feature maps exhibit similar behavior as observed for problems based only on shapes and textures present in images (Deepak and Ameer, 2020; Liu et al., 2018). In addition, the HS CNN-SVM also showed greater stability, when compared only to the use of CNN with reduced convergence time. Although the robustness of the model is one of the main targets in the development of a Computer Vision System, its stability is also important because it ensures that changes in the dataset will not affect the model's ability to adequately classify samples (Kaur and Kaur, 2015).
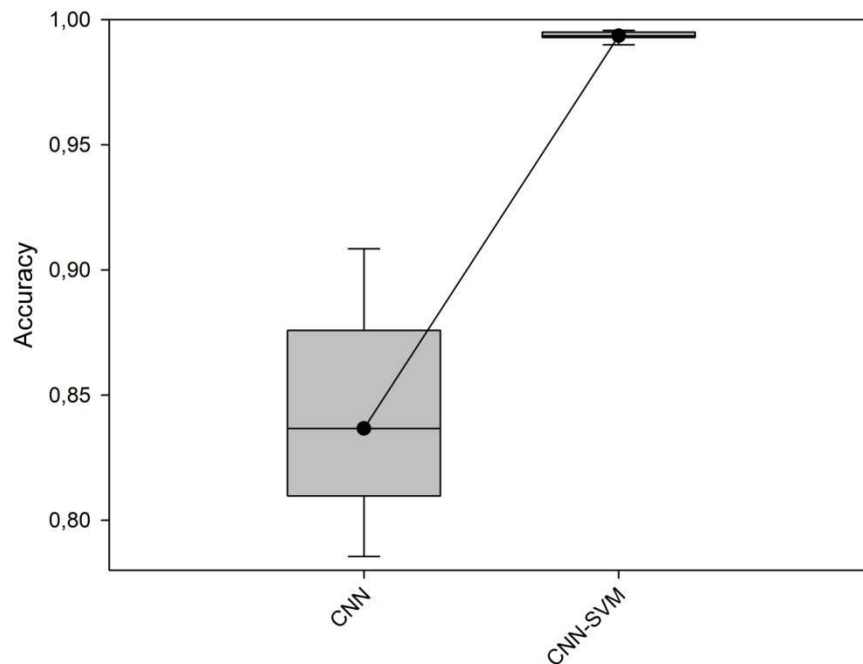
Figure 4. Overall accuracy of the CNN-SVM Hybrid System.

Table 1 shows the indicators individualized by class. The presentation of individualized indicators by class aims to show that HS CNN-SVM does not privilege any of the classes, despite the unbalance in the number of images among the classes present in the dataset (Goutte and Gaussier, 2005). As observed for accuracy, the HS CNN-SVM presents a correct prediction frequency higher than 99.50% for all classes under study. When evaluating the frequency of hits among the positive classifications, accuracy of 97.45% (±0.70%) for stage 3 was observed. The other stages presented accuracy higher than 99.20%. Recall measures the frequency of correct classification of the Hybrid System compared to the expected for each class (Kakhki et al., 2019). In this case, stage 4 presented an average recall of 97.39% (±0.73%) with the others presenting values higher than 99.00%. Although stages 3 and 4 presented precision and recall, respectively, lower than the other stages, the mean harmonic between these two indicators, given by the F1 Score, showed that the mean value was higher than 98.59%.

Table 1. Accuracy, precision, recall, F1 Score and Area Under the Curve (AUC), global and segmented by class, for the CNN-SVM Hybrid System, using the test dataset.

Stage	Accuracy		Precision		Recall		F1 Score		AUC	
	Mean	±SD	Mean	±SD	Mean	±SD	Mean	±SD	Mean	±SD
Stage 0	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000	1.0000	0.0000
Stage 1	0.9989	0.0007	1.0000	0.0000	0.9907	0.0059	0.9953	0.0030	1.0000	0.0000
Stage 2	0.9989	0.0007	0.9920	0.0050	1.0000	0.0000	0.9960	0.0025	1.0000	0.0000
Stage 3	0.9956	0.0013	0.9745	0.0070	0.9978	0.0029	0.9860	0.0041	0.9996	0.0001
Stage 4	0.9957	0.0013	0.9983	0.0027	0.9739	0.0074	0.9859	0.0042	0.9982	0.0018
Stage 5	0.9989	0.0009	0.9932	0.0054	0.9988	0.0026	0.9960	0.0033	1.0000	0.0000
Stage 6	0.9991	0.0007	0.9993	0.0023	0.9936	0.0053	0.9964	0.0029	1.0000	0.0000
Global	**0.9935**	**0.0018**	**0.9984**	**0.0007**	**0.9935**	**0.0018**	**0.9960**	**0.0011**	**0.9997**	**0.0003**

These results show that the color changes observed throughout the baking period, ranging from bright light to dark brown (Figure 1-C), are correctly recognized and classified by the HS CNN-SVM. Although the capacity of CNN-SVM hybrid systems to deal with less complex datasets was already known, such as the processing of Electroencephalography signals (EEG Signal) (Shi and Yang, 2019) and the processing of large RNA sequences for gene expression classification (Huynh et al., 2019), little was known about their capacity to deal with color changes. Now, it is evident that CNN-SVM hybrid systems can also be successfully used in problems involving color changes. It is also clear that the adoption of HS CNN-SVM, besides overcoming computational vision systems based only on CNN, also contributes to reduce the consumption of computational resources, such as memory and computational processing time, allowing its use in embedded or mobile systems. This opens up a range of application possibilities for the food industry, enabling the adoption of HS CNN-SVM in the control of processes that involve color changes, such as enzymatic and non-enzymatic browning, in pigment degradation. Besides these, it is also possible to envisage applications in the area of sensory analysis and recognition of ripening stages in fruits and vegetables.

5. Conclusions

In this wok, a CNN-SVM Hybrid System was proposed as a non-destructive tool for the recognition and classification of bread baking stages, based only on color changes in the crust of samples. The system combines the ability of CNN to extract the color feature map in images, without the need for human intervention, with the ability of SVM to discriminate classes. In the implementation of the Hybrid System, a CNN with reduced number of convolutional layers was adopted, followed by a SVM classifier, equipped with RBF kernel, using image fragments of the bread crust, throughout the baking period, as a training and testing dataset.

The obtained results allow the following conclusions to be drawn. First, the HS CNN-SVM is capable of recognizing and classifying the baking stages of bread samples, without human intervention, consisting of a fundamental tool for implementing an efficient control system. Second, the reduction in convergence time did not affect the HS CNN-SVM's ability to extract a feature map of colors present in the images, combining low computational processing time requirements during the training stage with the ability to classify samples based solely on the colors present in the bread crust. Third, the use of the hybrid system CNN-SVM allows a great reduction in memory consumption, allowing the use in mobile or on-board systems. Finally, the HS CNN-SVM proved to be superior to CNN-based classifiers for the same dataset. Future work will evaluate the effects of increasing the number of classes on the overall performance of the CNN-SVM HS on small datasets.

References

Azman, A.A., Ismail, F.S., 2017. Convolutional Neural Network for Optimal Pineapple Harvesting. Elektr. J. Electr. Eng. 16, 1–4. https://doi.org/10.11113/eletrika.v16n2.54

Battineni, G., Chintalapudi, N., Amenta, F., 2019. Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM). Informatics Med. Unlocked 16, 100200. https://doi.org/10.1016/j.imu.2019.100200

Bertrand, E., El Boustany, P., Faulds, C.B., Berdagué, J.-L., 2018. The Maillard Reaction in Food: An Introduction. Ref. Modul. Food Sci. 1–10. https://doi.org/10.1016/b978-0-08-100596-5.21459-5

Broyart, B., Trystram, G., 2003. Modelling of Heat and Mass Transfer Phenomena and Quality Changes During Continuous Biscuit Baking Using Both Deductive and Inductive (Neural Network) Modelling Principles. Food Bioprod. Process. 81, 316–326. https://doi.org/10.1205/096030803322756402

Chen, P.Y., Blutinger, J.D., Meijers, Y., Zheng, C., Grinspun, E., Lipson, H., 2019. Visual modeling of laser-induced dough browning. J. Food Eng. 243, 9–21. https://doi.org/10.1016/j.jfoodeng.2018.08.022

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297. https://doi.org/10.1007/BF00994018

Cotrim, W. da S., Minim, V.P.R., Felix, L.B., Minim, L.A., 2020. Short convolutional neural networks applied to the recognition of the browning stages of bread crust. J. Food Eng. 277, 109916. https://doi.org/10.1016/j.jfoodeng.2020.109916

Deepak, S., Ameer, P.M., 2020. Automated Categorization of Brain Tumor from MRI Using CNN features and SVM. J. Ambient Intell. Humaniz. Comput. https://doi.org/10.1007/s12652-020-02568-w

Gao, F., Wu, T., Li, J., Zheng, B., Ruan, L., Shang, D., Patel, B., 2018. SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis. Comput. Med. Imaging Graph. 70, 53–62. https://doi.org/10.1016/j.compmedimag.2018.09.004

Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. Proc. 13th Int. Conf. Artif. Intell. Stat. 9, 249–256.

Goutte, C., Gaussier, E., 2005. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. Lect. Notes Comput. Sci. 3408, 345–359. https://doi.org/10.1007/978-3-540-31865-1_25

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., Chen, T., 2018. Recent advances in convolutional neural networks.

Pattern Recognit. 77, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

Guo, Y., He, Y., Song, H., He, W., Yuan, K., 2018. Correlational examples for convolutional neural networks to detect small impurities. Neurocomputing 295, 127–141. https://doi.org/10.1016/j.neucom.2018.03.017

He, K., Sun, J., 2015. Convolutional neural networks at constrained time cost, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 5353–5360. https://doi.org/10.1109/CVPR.2015.7299173

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-Decem, 770–778. https://doi.org/10.1109/CVPR.2016.90

Helou, C., Jacolot, P., Niquet-Léridon, C., Gadonna-Widehem, P., Tessier, F.J., 2016. Maillard reaction products in bread: A novel semi-quantitative method for evaluating melanoidins in bread. Food Chem. 190, 904–911. https://doi.org/10.1016/j.foodchem.2015.06.032

Hoang, L., Lee, S., Kwon, K., 2020. A 3D Shape Recognition Method Using Hybrid Deep Learning Network CNN–SVM. Electronics 9, 649. https://doi.org/10.3390/electronics9040649

Hsiao, T.-Y., Chang, Y.-C., Chou, H.-H., Chiu, C.-T., 2019. Filter-based deep-compression with global average pooling for convolutional networks. J. Syst. Archit. 95, 9–18. https://doi.org/10.1016/j.sysarc.2019.02.008

Huynh, P., Nguyen, V., Do, T., 2019. Novel hybrid DCNN–SVM model for classifying RNA-sequencing gene expression data. J. Inf. Telecommun. 3, 533–547. https://doi.org/10.1080/24751839.2019.1660845

Ide, H., Kobayashi, T., Watanabe, K., Kurita, T., 2020. Robust pruning for efficient CNNs. Pattern Recognit. Lett. 135, 90–98. https://doi.org/10.1016/j.patrec.2020.03.034

Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, in: Bach, F., Blei, D. (Eds.), Proceedings of the 32 Nd International Conference on Machine Learning. Lille, pp. 448–456.

Kakhki, F.D., Freeman, S.A., Mosher, G.A., 2019. Evaluating machine learning performance in predicting injury severity in agribusiness industries. Saf. Sci. 117, 257–262. https://doi.org/10.1016/j.ssci.2019.04.026

Kaur, A., Kaur, K., 2015. An Empirical Study of Robustness and Stability of Machine Learning Classifiers in Software Defect Prediction, in: El-Alfy, E.-S.M., Thampi, S.M., Takagi, H., Piramuthu, S., Hanne, T. (Eds.), Advances in Intelligent Systems and Computing, Advances in Intelligent Systems and Computing.

Springer International Publishing, Cham, pp. 383–397.
https://doi.org/10.1007/978-3-319-11218-3_35

Kingma, D.P., Ba, J., 2015. Adam: A Method for Stochastic Optimization, in:
International Conference on Learning Representations (ICLR 2015). Ithaca, NY:
arXiv.org, San Diego, pp. 1–15.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation
and model selection, in: Proceedings of the 14th International Joint Conference
on Artificial Intelligence. pp. 1137–1143.
https://doi.org/10.1067/mod.2000.109032

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep
Convolutional Neural Networks, in: Advances in Neural Information Processing
Systems 25 (NIPS 2012). pp. 1–9. https://doi.org/10.1016/B978-008046518-
0.00119-7

Lathuilière, S., Mesejo, P., Alameda-pineda, X., Horaud, R., 2019. A Comprehensive
Analysis of Deep Regression. IEEE Trans. Pattern Anal. Mach. Intell. 41, 1–17.
https://doi.org/10.1109/TPAMI.2019.2910523

Li, H.X., Si, H., 2017. Control for Intelligent Manufacturing: A Multiscale Challenge.
Engineering 3, 608–615. https://doi.org/10.1016/J.ENG.2017.05.016

Lin, M., Chen, Q., Yan, S., 2014. Network In Network, in: 2nd International
Conference on Learning Representations, ICLR 2014 - Conference Track
Proceedings. Banff, pp. 1–10.

Lin, Y.-H., Chou, C.-N., Chang, E.Y., 2019. MBS: Macroblock Scaling for CNN Model
Reduction, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern
Recognition (CVPR). IEEE, pp. 9109–9117.
https://doi.org/10.1109/CVPR.2019.00933

Liu, B., Yu, X., Zhang, P., Yu, A., Fu, Q., Wei, X., 2018. Supervised Deep Feature
Extraction for Hyperspectral Image Classification. IEEE Trans. Geosci. Remote
Sens. 56, 1909–1921. https://doi.org/10.1109/TGRS.2017.2769673

Lym, S., Choukse, E., Zangeneh, S., Wen, W., Sanghavi, S., Erez, M., 2019.
PruneTrain: Fast Neural Network Training by Dynamic Sparse Model
Reconfiguration, in: Proceedings of the International Conference for High
Performance Computing, Networking, Storage and Analysis. ACM, New York,
NY, USA, pp. 1–13. https://doi.org/10.1145/3295500.3356156

Meng, Q., Chen, W., Wang, Y., Ma, Z.-M., Liu, T.-Y., 2019. Convergence analysis of
distributed stochastic gradient descent with shuffling. Neurocomputing 337, 46–
57. https://doi.org/10.1016/j.neucom.2019.01.037

Nadian, M.H., Rafiee, S., Aghbashlo, M., Hosseinpour, S., Mohtasebi, S.S., 2015.

Continuous real-time monitoring and neural network modeling of apple slices color changes during hot air drying. Food Bioprod. Process. 94, 263–274. https://doi.org/10.1016/j.fbp.2014.03.005

Nagpal, C., Dubey, S.R., 2019. A Performance Evaluation of Convolutional Neural Networks for Face Anti Spoofing, in: 2019 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 1–8. https://doi.org/10.1109/IJCNN.2019.8852422

Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines, in: Fürnkranz, J., Joachims, T. (Eds.), Proceedings of the 27th International Conference on Machine Learningence. Omnipress, Haifa, pp. 807–814.

Nalepa, J., Kawulok, M., 2019. Selecting training sets for support vector machines: a review. Artif. Intell. Rev. 52, 857–900. https://doi.org/10.1007/s10462-017-9611-1

Oh, H., Yu, Y., Ryu, G., Ahn, G., Jeong, Y., Park, Y., Seo, J., 2020. Convergence-Aware Neural Network Training, in: 2020 57th ACM/IEEE Design Automation Conference (DAC). IEEE, pp. 1–6. https://doi.org/10.1109/DAC18072.2020.9218518

Purlis, E., Salvadori, V.O., 2009. Modelling the browning of bread during baking. Food Res. Int. 42, 865–870. https://doi.org/10.1016/j.foodres.2009.03.007

Putranto, A., Chen, X.D., Zhou, W., 2015. Bread baking and its color kinetics modeled by the spatial reaction engineering approach (S-REA). Food Res. Int. 71, 58–67. https://doi.org/10.1016/j.foodres.2015.01.029

Rafegas, I., Vanrell, M., 2018. Color encoding in biologically-inspired convolutional neural networks. Vision Res. 151, 7–17. https://doi.org/10.1016/j.visres.2018.03.010

Razzaghi, T., Roderick, O., Safro, I., Marko, N., 2016. Multilevel weighted support vector machine for classification on healthcare data with missing values. PLoS One 11, 1–18. https://doi.org/10.1371/journal.pone.0155119

Santos, C.A., Welfer, D., 2019. A Novel Hybrid SVM-CNN Method for Extracting Characteristics and Classifying Cattle Branding. Lat. Am. J. Comput. VI, 9–16.

Schmittmann, O., Lammers, P., 2017. A true-color sensor and suitable evaluation algorithm for plant recognition. Sensors (Switzerland) 17. https://doi.org/10.3390/s17081823

Shen, Y., Zhou, H., Li, J., Jian, F., Jayas, D.S., 2018. Detection of stored-grain insects using deep learning. Comput. Electron. Agric. 145, 319–325. https://doi.org/10.1016/j.compag.2017.11.039

Shi, X., Yang, J., 2019. Hybrid System of Convolutional Neural Networks and SVM for Multi-class MIEEG Classification, in: 2019 IEEE 11th International Conference on Advanced Infocomm Technology (ICAIT). IEEE, pp. 137–141. https://doi.org/10.1109/ICAIT.2019.8935919

Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition, in: Bengio, Y., LeCun, Y. (Eds.), 3rd International Conference on Learning Representations, ICLR 2015. San Diego, pp. 1–14.

Tran, D.T., Iosifidis, A., Gabbouj, M., 2018. Improving efficiency in convolutional neural networks with multilinear filters. Neural Networks 105, 328–339. https://doi.org/10.1016/j.neunet.2018.05.017

Tseng, T.L. (Bill), Aleti, K.R., Hu, Z., Kwon, Y. (James), 2016. E-quality control: A support vector machines approach. J. Comput. Des. Eng. 3, 91–101. https://doi.org/10.1016/j.jcde.2015.06.010

Wang, J., Ma, Y., Zhang, L., Gao, R.X., Wu, D., 2018. Deep learning for smart manufacturing: Methods and applications. J. Manuf. Syst. 48, 144–156. https://doi.org/10.1016/j.jmsy.2018.01.003

Wu, C.W., 2018. ProdSumNet: reducing model parameters in deep neural networks via product-of-sums matrix decompositions.

Wu, H., Zhao, J., 2018. Deep convolutional neural network model based chemical process fault diagnosis. Comput. Chem. Eng. 115, 185–197. https://doi.org/10.1016/j.compchemeng.2018.04.009

Xu, Z., Yu, F., Qin, Z., Liu, C., Chen, X., 2020. DiReCtX: Dynamic Resource-Aware CNN Reconfiguration Framework for Real-Time Mobile Applications. IEEE Trans. Comput. Des. Integr. Circuits Syst. XX, 1–1. https://doi.org/10.1109/TCAD.2020.2995813

Yuan, X., Qi, S., Shardt, Y.A.W., Wang, Y., Yang, C., Gui, W., 2020. Soft sensor model for dynamic processes based on multichannel convolutional neural network. Chemom. Intell. Lab. Syst. 203, 104050. https://doi.org/10.1016/j.chemolab.2020.104050

Zanoni, B., Peri, C., Bruno, D., 1995. Modelling of browning kinetics of bread crust during baking. LWT - Food Sci. Technol. 28, 604–609. https://doi.org/10.1016/0023-6438(95)90008-X

Zhang, L., Putranto, A., Zhou, W., Boom, R.M., Schutyser, M.A.I., Chen, X.D., 2016. Miniature bread baking as a timesaving research approach and mathematical modeling of browning kinetics. Food Bioprod. Process. 100, 401–411. https://doi.org/10.1016/j.fbp.2016.08.007

Zhang, M.L., Zhou, Z.H., 2014. A Review on Multi-Label Learning Algorithms. IEEE

Trans. Knowl. Data Eng. 26, 1819–1837. https://doi.org/10.1109/TKDE.2013.39

Zhang, Q., Zhuo, L., Li, J., Zhang, J., Zhang, H., Li, X., 2018. Vehicle color recognition using Multiple-Layer Feature Representations of lightweight convolutional neural network. Signal Processing 147, 146–153. https://doi.org/10.1016/j.sigpro.2018.01.021

Zhang, X., Li, X., Feng, Y., Liu, Z., 2015. The use of ROC and AUC in the validation of objective image fusion evaluation metrics. Signal Processing 115, 38–48. https://doi.org/10.1016/j.sigpro.2015.03.007

Zhang, Y., Lian, J., Fan, M., Zheng, Y., 2018. Deep indicator for fine-grained classification of banana's ripening stages. Eurasip J. Image Video Process. 2018. https://doi.org/10.1186/s13640-018-0284-8

Zhong, R.Y., Xu, X., Klotz, E., Newman, S.T., 2017. Intelligent Manufacturing in the Context of Industry 4.0: A Review. Engineering 3, 616–630. https://doi.org/10.1016/J.ENG.2017.05.015

5. ARTIGO 4 - COMPUTER-ASSISTED COFFEE BEAN ROASTING CONTROL SYSTEM BASED ON DEEP LEARNING

Neste artigo será demonstrado o potencial de aplicação das redes neurais convolucionais para classificação dos estágios de torra de café baseado unicamente em imagens do processo capturadas com uma câmera fotográfica equipada com sensor CCD. Também será demonstrado que uma CNN pode ser treinada para estimar o tempo restante para que uma dada amostra atinja um ponto de torra arbitrário definido pelo operador, com base em imagens do processo. O manuscrito apresentado a seguir foi submetido para publicação no *Journal of Food Engineering*.

Computer-assisted coffee bean roasting control system based on deep learning

Weskley da Silva Cotrim[1], Leonardo Bonato Felix[2], Valéria Paula Rodrigues Minim[3], Renata Cássia Campos[4], Luis Antônio Minim[5,*]

[1] Universidade Federal de Mato Grosso, *Campus* Universitário do Araguaia, Instituto de Ciências Exatas e da Terra, Avenida Valdon Varjão, nº 6390. Barra do Garças - Mato Grosso, Brasil, CEP: 78600-000, weskleycotrim@ufmt.br, https://orcid.org/0000-0002-8190-9632

[2] Universidade Federal de Viçosa, Departamento de Engenharia Elétrica, *Campus* Universitário – Viçosa – Minas Gerais, Brasil, CEP: 36570-000, leobonato@ufv.br, https://orcid.org/0000-0002-6184-2354

[3] Universidade Federal de Viçosa, Departamento de Tecnologia de Alimentos, *Campus* Universitário – Viçosa – Minas Gerais, Brasil, CEP: 36570-000, vprm@ufv.br, https://orcid.org/0000-0001-7143-2060

[4] Universidade Federal de Viçosa, Departamento de Engenharia Agrícola, *Campus* Universitário – Viçosa – Minas Gerais, Brasil, CEP: 36570-000, renata@ufv.br, https://orcid.org/0000-0002-5419-5064

[5] Universidade Federal de Viçosa, Departamento de Tecnologia de Alimentos, *Campus* Universitário – Viçosa – Minas Gerais, Brasil, CEP: 36570-000, lminim@ufv.br, https://orcid.org/0000-0002-1584-9117

*Corresponding author
Prof. Luis Antônio Minim
Universidade Federal de Viçosa, Department of Food Technology, *Campus* Universitário – Viçosa – Minas Gerais, Brasil, CEP: 36570-000, lminim@ufv.br

ABSTRACT

Although the crop production and quality of coffee beans have been increasing every year, the control of the roasting process remains an eminently empirical activity, which results in high variability in the quality of the final beverage. In this sense, a computer-assisted coffee bean roasting control system was developed, based on Convolutional Neural Networks (CAS-CNN) and employing a low-cost CCD camera for image capture. The CAS-CNN was composed of a CNN with a reduced number of convolutional layers, trained with 864 images of coffee samples from the roasting process. In classification mode, the system presented an average global accuracy of 95.83% (±0.32%) and, in the roasting time prediction mode, it presented an average RMSE of 0.4 min (R^2=0.9870). The proposed CAS-CNN proved to be a powerful non-invasive and non-destructive tool, completely aligned with Industry 4.0, for monitoring and predicting coffee roasting time, ensuring the standardization and optimization of the process.

Keywords: coffee, roasting, deep learning, convolutional neural networks, computer vision, Industry 4.0.

Abbreviations
ANN - Artificial Neural Networks
CAS-CNN - computer-assisted control system based on Convolutional Neural Networks

CCD - Charge-Coupled Device
CIELab - Commission Internationale de l'Éclairage Color Space
CNN - Convolutional Neural Network
CVS - Computational Vision Systems
GAP - Global Average Pooling
GPU - Graphics Processing Unit
GRNN - General Regression Neural Network
MAE - Mean Absolute Error
NRMSE - Normalized Root Mean Square Error
RGB – Red, Green, Blue Color Space
ReLU - Rectified Linear Unit
RMSE - Root Mean Square Error
SD-CNN - Shallow-Deep Convolutional Neural Network
Short-CNN – Short Convolutional Neural Network
SVM - Support Vector Machine
sRGB - Standard RGB
VGGNet - Visual Geometry Group Neural Network

Conflict of interest
The authors have no conflicts of interest to declare that are relevant to the content of this article.

1. Introduction

Coffee is one of the most popular and consumed beverages on the planet (Czech et al., 2016). In the 2020/2021 crop year, coffee production was estimated at 175.5 million 60 kg bags and world consumption at around 165.4 million bags. Over the past five harvests, world coffee production increased by 8.5% and was accompanied by a 7.5% growth in consumption of the roasted coffee bean (USDA, 2020). In addition to the increase in production and consumption, there is also an increase in the quality of the coffee bean cultivated in the field, which reflects the increased demand for specialty coffees (Urwin et al., 2019). This increase in the quality of the coffee bean produced in the field has been obtained at the cost of the selection of varieties adapted for specific climates and soils, correct cultivation practices, and proper post-harvest management (Deribe, 2019). However, all the quality obtained during coffee bean cultivation and maintained during the post-harvest stage can be lost if the roasting stage is performed improperly (Garcia et al., 2018). This, because it is during the roasting step that the compounds responsible for the aroma, body, acidity, and characteristic color of the final beverage are formed (Baggenstoss et al., 2008; Buffo & Cardelli-Freire, 2004; X. Wang & Lim, 2017).

Among the physical-chemical changes that occur in the coffee bean during the roasting process, the color change is the most important, since it is used as an indicator to control the process and the final point of the operation (Buffo & Cardelli-

Freire, 2004). This is because the changes in the coffee bean color correlate with other important physical-chemical indicators in the roasting process. Among such indicators are the final moisture content of the bean, the volume increase, the mass loss (Bicho et al., 2012; Vargas-Elías et al., 2016; Virgen-Navarro et al., 2016), the content of volatile compounds produced during roasting (Baggenstoss et al., 2008; Kim et al., 2018; Sacchetti et al., 2009), the processing time (Romani et al., 2012) and the content of carbon dioxide formed (X. Wang & Lim, 2017).

The measurement of color change in coffee beans can be performed indirectly, by determining the content of compounds derived from the Maillard reaction (A. Echavarría et al., 2014; Moreira et al., 2012), or directly, by instrumental colorimetric analysis (Kim et al., 2018). However, the most common is the monitoring of color changes performed by visual inspection of the beans, throughout the roasting process, by a trained operator. The visual inspection technique, using colored standards for comparison, is considered to be lower cost, however, it presents low accuracy, high variability in results, and significant dependence on a highly specialized operator, which can compromise the final quality of the roasted coffee (Garcia et al., 2018). On the other hand, the use of instrumental techniques, such as spectrophotometry, on the Agtron scale, or colorimetry, on the CIELab scale, represent high accuracy and reproducibility of results, but at a high cost of equipment acquisition and maintenance, which makes them prohibitive for most companies (Leme et al., 2019; Schmittmann & Lammers, 2017). Additionally, instrumental techniques usually require the coffee sample to be ground before color determination, which limits their use in real-time (Craig et al., 2018). And finally, the adoption of indirect techniques, as is the case with the determination of melanoidin content, presents intermediate cost, but demands long analysis times to obtain the results (A. P. Echavarría et al., 2012).

Due to these characteristics, such techniques become unsuitable for use in industrial control systems, especially in the context of Industry 4.0. The reason for this is that the 4th generation industry is based on the use of intelligent systems and equipment, in which information must be easily obtained and shared autonomously, allowing rapid decision making, minimizing losses, and reducing costs (Li & Si, 2017; J. Wang et al., 2018; Zhong et al., 2017). Thus, the search for rapid and low-cost methods for measuring and tracking color changes in coffee during roasting are goals to be achieved. To meet the assumptions of Industry 4.0, the proposed technique should be based on objective and non-destructive measurements, which can be applied in the process line and does not depend on human intervention for sample or information preprocessing.

In this context, the adoption of Artificial Intelligence models, especially Convolutional Neural Networks (CNN), emerge as possible candidates for building computer-assisted coffee bean roasting control systems. CNN are known for their ability to recognize complex patterns present in images and for dealing well with

nonlinearities (He et al., 2016; Krizhevsky et al., 2012; Simonyan & Zisserman, 2015), such as those present in food processing.

The first CNN was proposed by Le Cun and collaborators for handwritten character recognition (Le Cun et al., 1989). The great novelty lay in the ability of the neural network, trained using the back-propagation algorithm, to recognize and extract complex patterns present in large image datasets without the need for human intervention (Lecun et al., 1998). Such neural network became known as LeNet-5 and is considered a milestone for the development of deep learning (Gu et al., 2018; H. Wu & Zhao, 2018).

Although the LeNet-5 architecture was considered revolutionary, the computational cost for its implementation limited its practical use. Only in 2012, with the advance of GPU (Graphics Processing Unit) computing, a new architecture emerged, known as AlexNet (Krizhevsky et al., 2012), reactivating the interest of researchers in convolutional neural networks. Over the years, several architectures have been proposed, always focusing on increasing the performance of the neural network, most notably VGGNet (Simonyan & Zisserman, 2015), a GoogLeNet (Szegedy et al., 2015), and ResNet (He et al., 2016). However, increased performance has come at the price of increased computational cost, which has led many researchers to propose shallower CNN for use in mobile equipment and embedded systems, such as Short-CNN (Cotrim et al., 2020), SD-CNN (Gao et al., 2018) or even hybrid systems CNN-SVM (Cotrim, Felix, et al., 2021).

The CNN have been widely used in building Computer Vision Systems (CVS) for detecting and classifying objects present in images, always based on their texture and edge features. In industrial processes, CNN have been applied in the detection of insects present in stored grains (Shen et al., 2018), in the diagnosis and detection of faults in chemical processes (H. Wu & Zhao, 2018), and in the detection of small impurities in bottled beverages (Guo et al., 2018). Meanwhile, recent studies have demonstrated the potential of using CNN in solving problems involving color, as is the case for ripeness stage classification of pineapple (Azman & Ismail, 2017) and banana (Yan Zhang et al., 2018). It has also been shown that CNN are effectively able to extract color features present in images and use them for classification (Rafegas & Vanrell, 2018). More recently, the potential for using CNN to solve problems involving color changes has been demonstrated (Cotrim et al., 2020).

Although CNN have primary application in image classification, recent studies have also demonstrated the potential for using these networks to solve regression problems that involve continuous numerical estimations. A good example is the application of CNN for estimating the remaining lifetime of systems based on sensor-provided data (Babu et al., 2016). The same technique was also used in automatic segmentation of the coronary lumen in images obtained by intravascular optical coherence tomography (Yong et al., 2017). Another important application of CNN in problems involving continuous data is the detection of multiple lanes on highways

(Chougule et al., 2019). Recent applications of this technique involve age estimation of people based on photos of their faces (Dornaika et al., 2020; H. Liu et al., 2019) and virtual sensor modeling for dynamic hydrocracking processes in the petrochemical industry (Yuan et al., 2020). Such applications highlight the potential for using CNN in the time estimation of industrial processes based on visual modifications, such as those observed in coffee roasting.

Thus, the objective of this work was to develop a computer vision system for computer-assisted control of coffee bean roasting using Convolutional Neural Network (CAS-CNN). Since traditional systems based on sensors operating in the CIELab color space or even by spectrophotometry are expensive, this work used a digital camera equipped with a Charge-Coupled Device (CCD) sensor for capturing images, which have low acquisition and maintenance costs. This CAS-CNN represents a great technological leap for roasted coffee producers, since it allows the standardization of roasting processes at low cost, with the additional advantage of being a technology compatible with Industry 4.0.

2. Material and Methods

2.1. Sample Preparation

Coffee beans (*Coffea arabica* L.), purchased directly from coffee producers in the state of Rio de Janeiro, Brazil, were used in the study. The raw coffee samples were roasted in an electrically heated drum roaster, equipped with an automatic temperature control system. A medium-speed roasting profile was adopted, with the initial temperature set to 190 °C and time for each roasting run varying from 2 to 12 min, with a time step of 2 min until first crack (8 min) and 1 min until the final time (12 min) (Figure 1c). Each roasting run was performed by adding 250 g of raw coffee beans to the roaster at a previously adjusted temperature. This whole process was carried out with three repetitions. After a pre-determined stopping time, the samples were collected, cooled, and sent to the image capture equipment.

2.2. Image Capture Device

An image capture device was built for the study, composed of an illumination chamber, a digital camera connected to a microcomputer, and a glass container to support the coffee samples (Figure 1a). The illumination chamber was built with dimensions of 250 x 200 x 250 mm (width x height x depth) and was equipped with an 8 W fluorescent lamp (6500 K). A digital camera (Canon EOS DIGITAL REBEL XT) was used, positioned in an opening at the top of the illumination chamber, 250 mm from the sample location, secured by a support screw. A glass container of 100 mm diameter and 20 mm height was used to position the samples inside the illumination chamber.
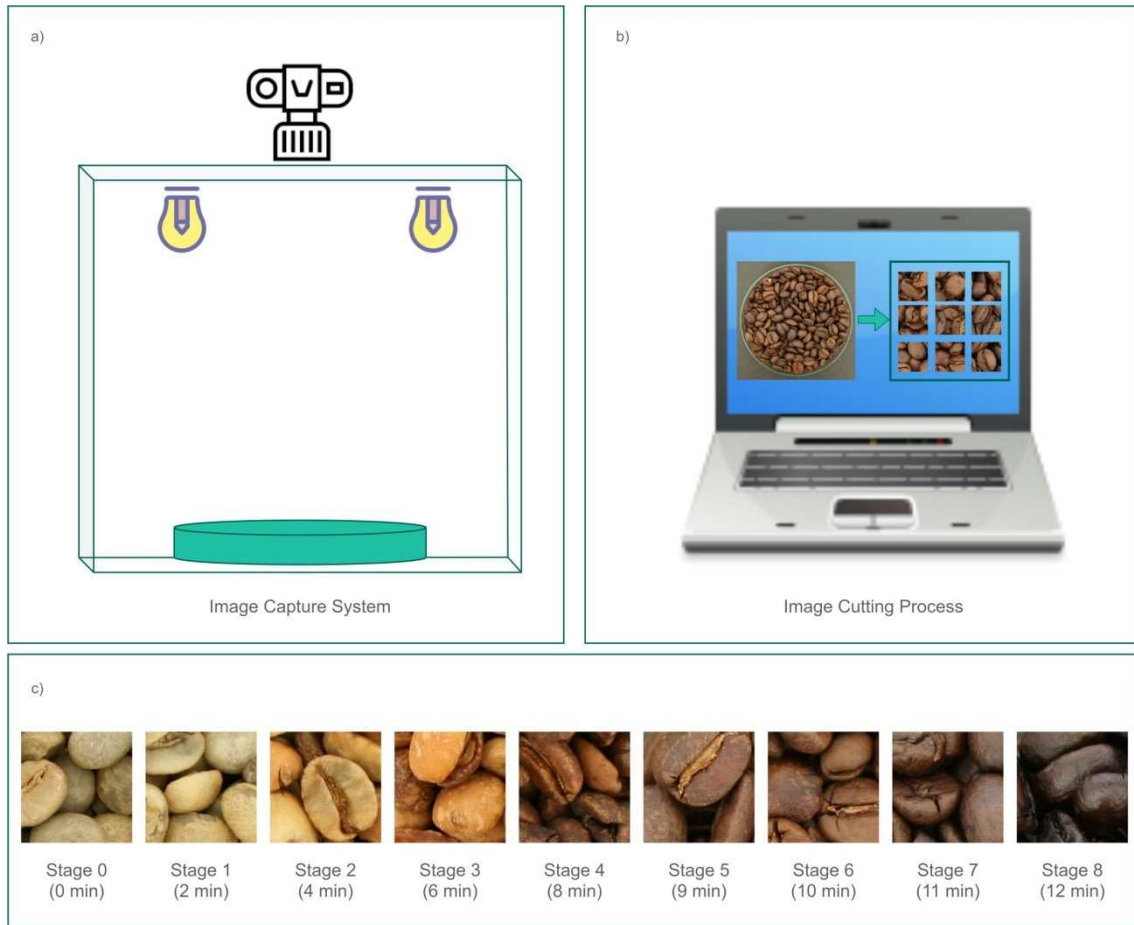
Figure 1. Graphic representation of a) the image capture device, b) the image cutting and processing system, and c) the different coffee roasting stages.

2.3. Acquisition of images

At the end of each roasting run, the sample was collected from the roaster in sufficient quantities to fill the glass container used as a sample holder. The sample container was positioned centrally inside the illumination chamber. The camera used to capture the images was remotely operated from the microcomputer by proprietary software (Digital Photo Professional Ver.3.10). The camera was set to operate with a focal length of 25 mm, ratio f/5, shutter speed of 1/50 sec, and ISO 100. Images were captured with color representation in sRGB (Standard RGB) space, with a resolution of 314 dpi and dimensions 3456 x 2304 pixels. After capture, the images were immediately transferred to the microcomputer by USB 2.0 data cable and proceeded to the pre-processing step.

2.4. Preprocessing of images and data augmentation

Each image was automatically cut with no overlapping, into fragments of 100 x 100 pixels, and the regions that did not contain exclusively coffee beans were discarded (Figure 1b). Thus, the dataset was composed of 864 unique images, distributed in 9 stages (Figure 1c). The images were then randomly subdivided into three subgroups: training (80%), validation (10%), and test (10%). To minimize the risk of overfitting, the Data Augmentation technique was applied to the dataset. Data

Augmentation was performed through the combined use of the geometric transformation techniques of translation, zooming, mirroring, and image rotation (-180° to 180°, variation of 1°). Data Augmentation factor equal to 7 was applied, a lower value than that used in similar works (Chen et al., 2019; Cotrim et al., 2020; Lathuiliere et al., 2020). After applying the Data Augmentation technique, the dataset increased to 6048 images. To meet the requirements of the adopted CNN architecture, the images were resized to 300 x 300 pixels size.

2.5. Architecture of convolutional neural network

A short CNN architecture described by Cotrim et al. (2021) was adopted, with modifications (Figure 2a). The CNN was composed of three convolutional layers (5x5, 3x3, and 3x3), with the first convolutional layer receiving images with dimensions 300 x 300 pixels. Pooling layers (3x3) were added immediately after layers 1 and 2, and all convolutional layers were normalized by Batch Normalization technique.
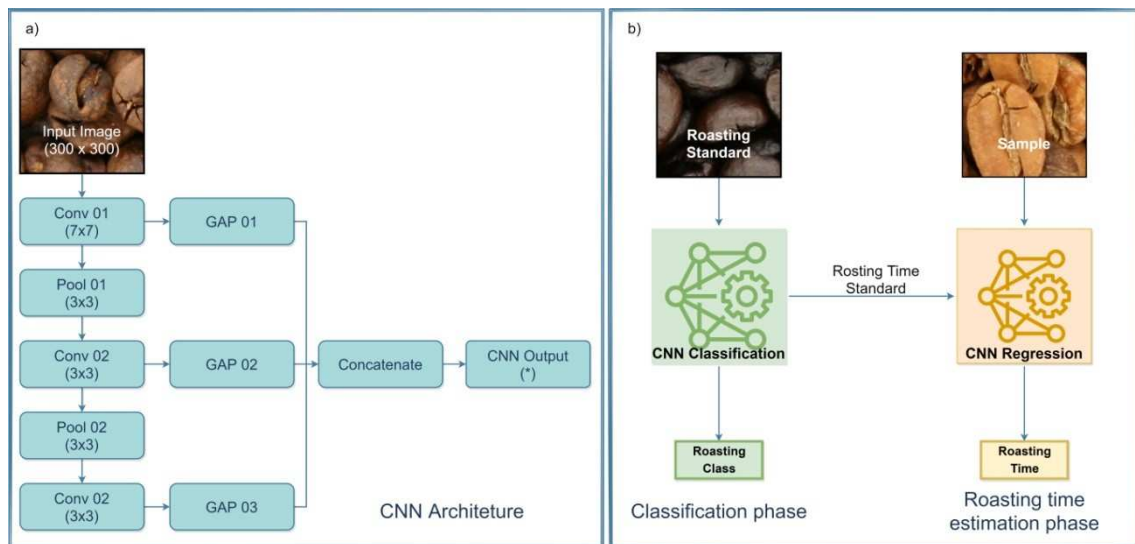

Figure 2. a) Architecture of the convolutional neural network and b) Classification phases of the roasting stages and estimation of the time required to reach a roasting pattern previously defined by the operator.
*Output layer in classification mode with Softmax activation function and nine roasting stages; Output layer in regression mode with ReLU activation function and a continuous variable (time).

A Global Average Pooling (GAP) function was applied to the output of each of the convolutional layers. The outputs of the three GAP functions were subsequently merged by a concatenation function. In classification mode, the output of the concatenation function links directly to the classification output layer, with a Softmax function and nine possible roasting stages. In regression mode, for estimating the time required to achieve an arbitrary roasting pattern set by an operator, the output layer relied on a ReLU activation function for a continuous output variable (time) (Figure 2b).

2.6. Training Strategy

CNN hyperparameters were adjusted using the Adam Optimizer algorithm (Kingma & Ba, 2015), with a batch size of 12 images, a learning rate of $1.0\text{x}10^{-3}$, and a learning rate decay rate equal to $1.0\text{x}10^{-3}$. The Glorot Uniform technique was used for the initialization of the convolutional kernels (Glorot & Bengio, 2010). Also, shuffling of the dataset was performed before each training epoch, to ensure that the feature map extraction was not influenced by the order of appearance of the images in the dataset (Ioffe & Szegedy, 2015; Meng et al., 2019). In the classification mode, the "Categorical Cross-Entropy Loss" function was adopted as the loss function to be minimized during the CNN training process. In regression mode, the "Mean Absolute Error" (MAE) function was chosen as the loss function to be minimized. The entire operation of image preprocessing, data augmentation, training, validation, and testing of the CNN was implemented by means of scripts written in Python 3.6, using the libraries TensorFlow 1.12, Keras 2.2, OpenCV 4.0, Matplotlib 2.2, and Numpy 1.16, in a microcomputer equipped with an Intel Core i5 6500 processor, 24 GB RAM, and an Nvidia GeForce RTX 2060 video card.

2.7. Model evaluation metrics

In the classification mode, the performance of the CNN was evaluated both globally and for each of the nine coffee roasting stages by using the indicators of accuracy, precision, revocation, and F1 score (X. Zhang et al., 2015). To ensure model consistency, the model was trained five times on the training dataset, always performing dataset randomization. The average of the indicators was obtained by applying the micro-average technique (M. L. Zhang & Zhou, 2014), according to Equation 1.

$$\left.\begin{aligned} Accuracy &= \frac{TP_j + TN_j}{TP_j + FP_j + TN_j + FN_j} \\ Precision &= \frac{TP_j}{TP_j + FP_j} \\ Recall &= \frac{TP_j}{TP_j + FN_j} \\ F1 &= \frac{2TP_j}{2TP_j + FP_j + FN_j} \end{aligned}\right\} (Eq.\,1)$$

where TP_j (True Positive) is the amount of images of a class under evaluation correctly classified in the *j-th* repetition, TN_j (True Negative) the amount of images not belonging to the class under evaluation correctly classified in the *j-th* repetition, FP_j (False Positive) the amount of images from another class incorrectly classified in the class under evaluation in the *j-th* repetition, and FN_j (False Negative) the amount of images from the class under evaluation incorrectly classified in another class in the *j-th* repetition.

The performance of the CNN in regression mode was evaluated by the overall and per-process time Root Mean Square Error (RMSE) and the overall Normalized Root Mean Square Error (NRMSE) (Cotrim, Coimbra, et al., 2021), according to Equation 2.

$$\left. \begin{array}{l} RMSE = \sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2} \\[2em] NRMSE(\%) = 100\left(\dfrac{RMSE}{y_{max} - y_{min}}\right) \end{array} \right\} (Eq.\,2)$$

where n is the number of observations in the dataset, y_i is the time (min) measured during the roasting process, $\hat{y}$ the time (min) estimated by the model, y_{max} the final roasting time (12 min) and y_{min} the initial roasting time (0 min).

3. Results and Discussion

A low-cost computer-assisted coffee roasting control system based on deep learning techniques was proposed, using only images captured by a camera with a CCD sensor throughout the process. To achieve the objectives, two basic premises were established: i) the adopted model should be able to extract all the necessary information for the classification of the roasting stages directly from the images captured during the process, without the need for human intervention, and ii) the computational resources requirement for the resulting system should be low, allowing it to operate in mobile equipment or embedded systems. In this sense, a Convolutional Neural Network (CNN) with a reduced number of convolutional layers was adopted, since it has already been successfully employed in solving problems involving colors (Cotrim, Felix, et al., 2021). Such architecture has only 322,857 parameters, a value considered low when compared to traditional CNN architectures, such as AlexNet that has 62 million parameters, or VGG16 that reaches 132 million parameters (Hsiao et al., 2019). A reduced number of trainable parameters contributes to a reduction in memory consumption and a reduction in the time required during the training step, and also contributes to a reduction in the risk of data memorization (C. W. Wu, 2018; Q. Zhang et al., 2018). Another advantage of using architectures with a reduced number of convolutional layers concerns the time required for processing each image after training the network. In CAS-CNN the processing time for a single image was 3 ms, which makes the process almost instantaneous and makes its real-time application feasible.

The adoption of a CNN with a reduced number of convolutional layers did not affect the performance of the CAS-CNN. The resulting system showed good discriminative ability of roasting stages, with a global average accuracy of 95.83% (±0.32%) obtained on a test dataset, different from that used in the training step (Figure 3). The model was correct (overall precision) 96.64% (±0.29%) of the times it assigned a sample to a given roasting stage, resulting in a low false-positive rate. In

addition, the model correctly classified (overall recall) 95.50% (±1.35%) of the samples tested. The overall reliability of the model was evaluated by the F1 Score indicator, which presented an overall average of 96.11% (±0.11%). Similar results were observed when using CNN with a reduced number of convolutional layers, which were successfully used in identifying eight vehicle colors by traffic cameras (Q. Zhang et al., 2018) and in identifying seven stages of bread baking (Cotrim et al., 2020). As seen, the adoption of CNN with a reduced number of convolutional layers does not compromise its discriminative ability in solving problems with a small number of classes. This characteristic guarantees low computational resource consumption, a high image processing rate, and the possibility of application in systems with limited hardware. Such information is very relevant for the food industry, where many problems present a small number of classes when compared with traditional computer vision problems in other areas (Azman & Ismail, 2017; Pfisterer et al., 2018; Yiheng Zhang et al., 2018).

In addition to the global indicators, the performance of the CNN segmented by roasting stage was also evaluated. The model showed homogeneous results of average accuracy per roasting stage, ranging from 96.01% (±0.22%) at stage 6 to 100.00% at stages 2 and 8. All roasting stages showed precision above 90%, except for stage 6, which had an average precision of 81.97% (±2.31%). Similar behavior was observed for the recall indicator, where only stages 5 and 6 showed averages lower than 90%, which was reflected in the average values of the F1 Score attributed to these two stages, respectively 85.99% (±1.68%) and 82.20% (±0.87%). The results observed for stages 5 and 6 may be associated with the reduction in the time interval between the stages. Between stages 0 and 5 a sample collection time interval of 2 min was adopted. From stage 5 onwards, the sampling interval was reduced to 1 min, since at this stage color changes occur at a higher rate (Figure 1c). A shorter interval between two stages leads to a greater similarity between the colors, potentially making it more difficult to correctly identify them. Added to this is the fact that experimentally, it was demonstrated that CNN deals better with the colors black, yellow, and red than with the gray color (Fu et al., 2020; Q. Zhang et al., 2018). This explains the increase in accuracy, recall, and F1 score values that occurred as the sample became darker in the phase when the sampling interval was reduced to 1 min.
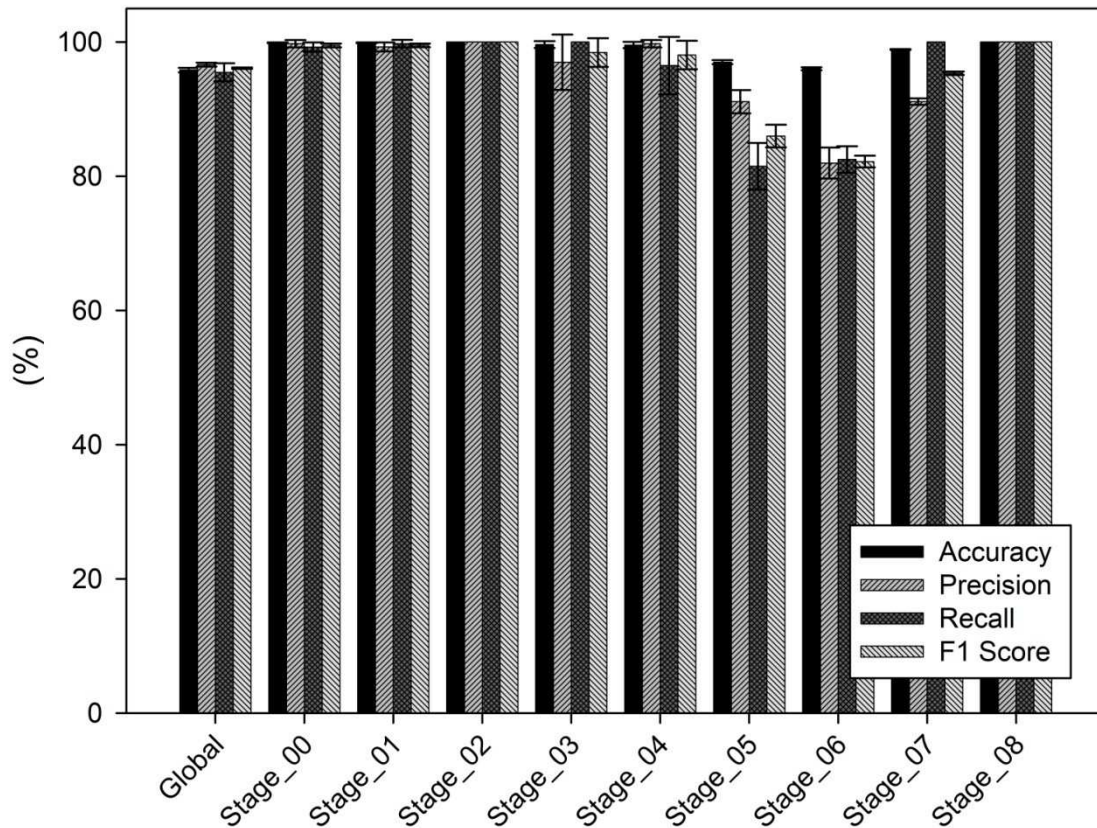
Figure 3. Accuracy, precision, recall, and F1 Score, global and stage-segmented, for the convolutional neural network (CAS-CNN) using the test dataset.

The CNN was also trained, in regression mode, to recognize in an image the current roasting stage and estimate the time required to reach the roasting standard arbitrarily stipulated by the operator, in the form of an image of the desired roasting stage. In this work, arbitrary images from stage 8 were adopted as the roasting standard, which corresponds to the 12-minute roasting time. The CAS-CNN presented a Normalized Root Mean Square Error (NRMSE) of 3,3% when adopted the test dataset for coffee roasting time prediction. In a similar coffee roasting time prediction problem, models based on artificial neural networks (ANN) presented an NRMSE of 9,5% and when using the general regression neural network (GRNN) the NRMSE was equal to 8,9% (Romani et al., 2012). These results show the superiority of the deep learning-based model (CAS-CNN) over those based on traditional machine learning techniques (ANN and GRNN) in recognizing and predicting coffee roasting time.

Figure 4 shows the relationship of predicted versus experimental values of the time required for the end of roasting (R^2=0.9870). The largest prediction errors were observed for the initial stages of the roasting process. As the roasting process proceeds and approaches its end, an increase in model predictive ability is observed. Such discrepancy in predictive ability seems to be a characteristic of this type of model, that is, the less the difference between the current process time and the desired final time, the lesser the error (H. Liu et al., 2019).
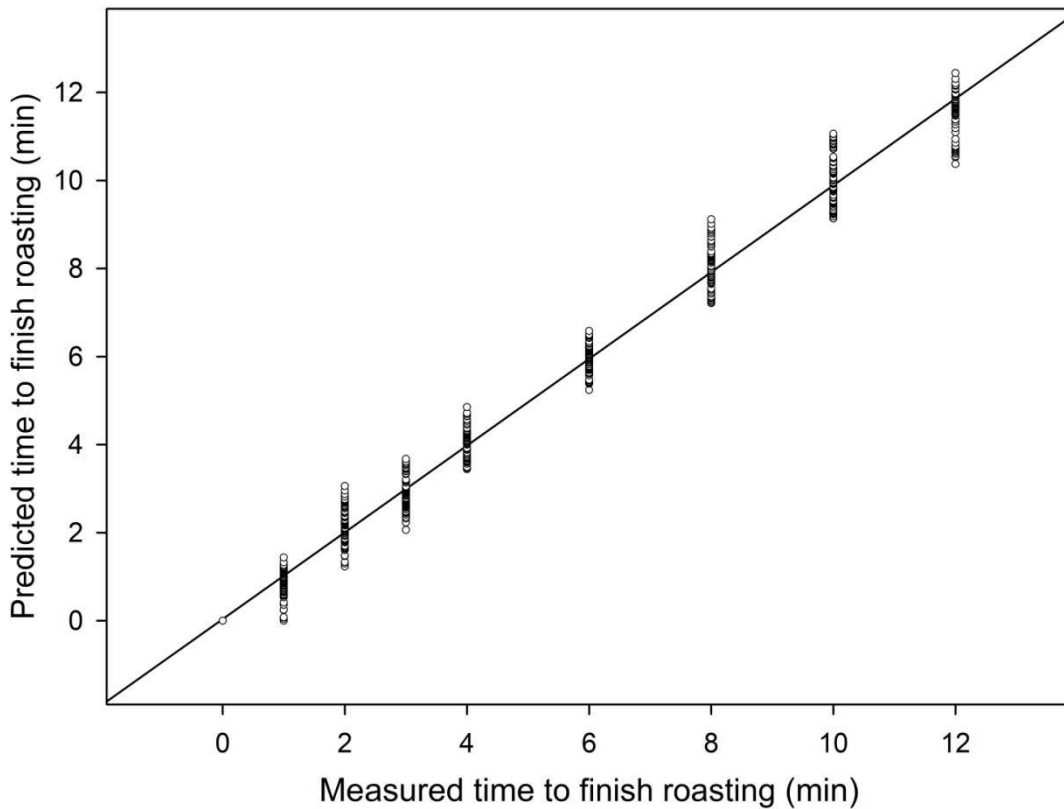
Figure 4. Predicted versus experimental values for the time required until the end of the roasting process, estimated by CAS-CNN on the test dataset.

This hypothesis is supported by the values of the Root Mean Square Error (RMSE) segmented by each of the times until the end of roasting (Figure 5). The highest RMSE value (0.7 min) was found for the time furthest from the roasting pattern (12 min), which points to the model's greater difficulty in estimating roasting time when the coffee bean is still raw. As the roasting process proceeds, a decrease in the RMSE value is observed, which reaches the lowest value (0.3 min) at the current roasting time of 6 min. After this time, an increase in the RMSE value is again observed. However, this increase in RMSE values is because, after the first crack, the image collection interval was reduced from 2 min to 1 min, which represents an increase in the complexity of the analysis for the model. However, as can be seen, this increase remains below the average value of the Root Mean Square Error (0.4 min), except for the stage immediately before (1 min) the roasting endpoint, which showed an RMSE of 0.5 min.
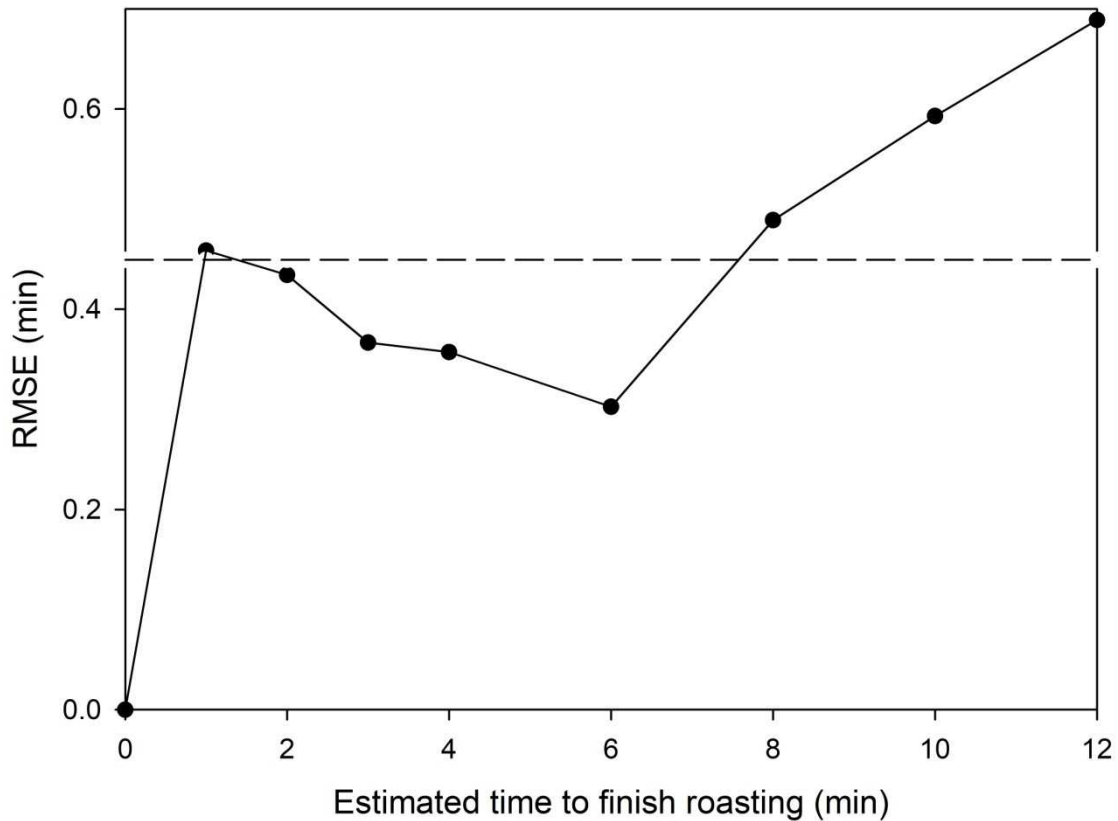
Figure 5. Root Mean Square Error (RMSE) segmented by the estimated time required to reach the end of roasting predicted by CAS-CN on the test dataset.

Finally, two points should be highlighted about the potential of using CNNs in building computer vision systems for coffee roasting control. The first, evidently concerns the superiority of CNNs in process time estimation over traditional models in nonlinear processes (Babu et al., 2016). During coffee roasting, a series of chemical reactions occur in parallel and in sequence forming the colored compounds responsible for the browning of the coffee bean (Madihah et al., 2012; Tsai et al., 2017). Such reactions occur in a non-linear and complex manner, which makes it difficult to adopt conventional models in their measurement and use for process time estimation (Santos et al., 2016; Virgen-Navarro et al., 2016; X. Wang & Lim, 2014). Another point to be highlighted refers to the current revolution experienced by industry (Industry 4.0), which has radically changed the operating logic of a factory. In the context of Industry 4.0, equipment and systems must be able to collect, process and share information autonomously (Zhong et al., 2017). In this sense, CNNs are perfectly suited since they do not rely on human intervention for the extraction of the features that will be used in the training and operation of the model (Y. Liu et al., 2021; J. Wang et al., 2018). Thus, the combination of CNN's ability to handle process nonlinearities with the ability to extract features directly from images makes this model an ideal tool for building computer vision systems for controlling and monitoring the coffee roasting process.

5. Conclusions

In this work, a computer-aided coffee roasting control system based on convolutional neural networks (CAS-CNN) was proposed. For the system implementation, a CNN with a reduced number of convolutional layers was adopted. Images of nine stages of coffee bean roasting were used as input data. The results obtained support the following conclusions. First, the use of our CNN, with a reduced number of convolutional layers, does not compromise the performance of the system in solving common food industry problems where the number of classes is low. Second, the CAS-CNN was shown to be superior to traditional methods for coffee roasting time prediction. Finally, the resulting system is a powerful non-invasive, and non-destructive tool for monitoring and predicting coffee roasting time, ensuring process standardization and optimization. Such abilities match perfectly with the fourth industrial revolution, currently experienced by the food industry, where systems and equipment must be able to collect, process, and share information autonomously, allowing rapid decision making. The results presented here are encouraging and point to the need for further studies to improve the system for operation with shorter time intervals for image collection or even using video capture, thus ensuring its application in real-time control systems.

References

Azman, A. A., & Ismail, F. S. (2017). Convolutional Neural Network for Optimal Pineapple Harvesting. *ELEKTRIKA- Journal of Electrical Engineering*, *16*(2), 1–4. https://doi.org/10.11113/eletrika.v16n2.54

Babu, G. S., Zhao, P., & Li, X.-L. (2016). Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life. In S. B. Navathe, W. Wu, S. Shekhar, X. Du, X. S. Wang, & H. Xiong (Eds.), *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 9642, pp. 214–228). Springer International Publishing. https://doi.org/10.1007/978-3-319-32025-0_14

Baggenstoss, J., Poisson, L., Kaegi, R., Perren, R., & Escher, F. (2008). Coffee Roasting and Aroma Formation: Application of Different Time−Temperature Conditions. *Journal of Agricultural and Food Chemistry*, *56*(14), 5836–5846. https://doi.org/10.1021/jf800327j

Bicho, N. C., Leitão, A. E., Ramalho, J. C., & Lidon, F. C. (2012). Use of colour parameters for roasted coffee assessment. *Food Science and Technology*, *32*(3), 436–442. https://doi.org/10.1590/s0101-20612012005000068

Buffo, R. A., & Cardelli-Freire, C. (2004). Coffee flavour: An overview. *Flavour and Fragrance Journal*, *19*(2), 99–104. https://doi.org/10.1002/ffj.1325

Chen, P. Y., Blutinger, J. D., Meijers, Y., Zheng, C., Grinspun, E., & Lipson, H. (2019). Visual modeling of laser-induced dough browning. *Journal of Food*

Engineering, *243*(April 2018), 9–21.
https://doi.org/10.1016/j.jfoodeng.2018.08.022

Chougule, S., Koznek, N., Ismail, A., Adam, G., Narayan, V., & Schulze, M. (2019). Reliable multilane detection and classification by utilizing CNN as a regression network. In L. Leal-Taixé & S. Roth (Eds.), *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 11133 LNCS*. Springer International Publishing. https://doi.org/10.1007/978-3-030-11021-5_46

Cotrim, W. da S., Coimbra, J. C., & Cotrim, K. C. F. (2021). Modeling and simulation of broiler carcass precooling by computational fluid dynamics. *Journal of Food Process Engineering*, *44*(6). https://doi.org/10.1111/jfpe.13693

Cotrim, W. da S., Felix, L. B., Minim, V. P. R., Campos, R. C., & Minim, L. A. (2021). Development of a hybrid system based on convolutional neural networks and support vector machines for recognition and tracking color changes in food during thermal processing. *Chemical Engineering Science*, *240*, 116679. https://doi.org/10.1016/j.ces.2021.116679

Cotrim, W. da S., Minim, V. P. R., Felix, L. B., & Minim, L. A. (2020). Short convolutional neural networks applied to the recognition of the browning stages of bread crust. *Journal of Food Engineering*, *277*, 109916. https://doi.org/10.1016/j.jfoodeng.2020.109916

Craig, A. P., Botelho, B. G., Oliveira, L. S., & Franca, A. S. (2018). Mid infrared spectroscopy and chemometrics as tools for the classification of roasted coffees by cup quality. *Food Chemistry*, *245*(November 2017), 1052–1061. https://doi.org/10.1016/j.foodchem.2017.11.066

Czech, H., Schepler, C., Klingbeil, S., Ehlert, S., Howell, J., & Zimmermann, R. (2016). Resolving Coffee Roasting-Degree Phases Based on the Analysis of Volatile Compounds in the Roasting Off-Gas by Photoionization Time-of-Flight Mass Spectrometry (PI-TOFMS) and Statistical Data Analysis: Toward a PI-TOFMS Roasting Model. *Journal of Agricultural and Food Chemistry*, *64*(25), 5223–5231. https://doi.org/10.1021/acs.jafc.6b01683

Deribe, H. (2019). Review on Factors which Affect Coffee (Coffea Arabica L.) Quality in South Western, Ethiopia. *International Journal of Forestry and Horticulture*, *5*(1), 12–19. https://doi.org/10.20431/2454-9487.0501003

Dornaika, F., Bekhouche, S., & Arganda-Carreras, I. (2020). Robust regression with deep CNNs for facial age estimation: An empirical study. *Expert Systems with Applications*, *141*, 112942. https://doi.org/10.1016/j.eswa.2019.112942

Echavarría, A. P., Pagán, J., & Ibarz, A. (2012). Melanoidins Formed by Maillard Reaction in Food and Their Biological Activity. *Food Engineering Reviews*, *4*(4),

203–223. https://doi.org/10.1007/s12393-012-9057-9

Echavarría, A., Pagán, J., & Ibarz, A. (2014). Kinetics of color development of melanoidins formed from fructose/amino acid model systems. *Food Science and Technology International*, *20*(2), 119–126. https://doi.org/10.1177/1082013213476071

Fu, H., Ma, H., Wang, G., Zhang, X., & Zhang, Y. (2020). MCFF-CNN: Multiscale comprehensive feature fusion convolutional neural network for vehicle color recognition based on residual learning. *Neurocomputing*, *395*, 178–187. https://doi.org/10.1016/j.neucom.2018.02.111

Gao, F., Wu, T., Li, J., Zheng, B., Ruan, L., Shang, D., & Patel, B. (2018). SD-CNN: A shallow-deep CNN for improved breast cancer diagnosis. *Computerized Medical Imaging and Graphics*, *70*, 53–62. https://doi.org/10.1016/j.compmedimag.2018.09.004

Garcia, C. D. C., Pereira Netto, A. D., Da Silva, M. C., Catão, A. A., De Souza, I. A., Farias, L. S., Emerich de Paula, T. N., Emerick de Paula, M. N., Dos Reis, S. C., & Da Silva Junior, A. I. (2018). Relative importance and interaction of roasting variables in coffee roasting process. *Coffee Science*, *13*(3), 379–388. https://doi.org/10.25186/cs.v13i3.1483

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the 13th International Conference On Artificial Intelligence and Statistics*, *9*, 249–256.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, *77*, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

Guo, Y., He, Y., Song, H., He, W., & Yuan, K. (2018). Correlational examples for convolutional neural networks to detect small impurities. *Neurocomputing*, *295*, 127–141. https://doi.org/10.1016/j.neucom.2018.03.017

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Hsiao, T.-Y., Chang, Y.-C., Chou, H.-H., & Chiu, C.-T. (2019). Filter-based deep-compression with global average pooling for convolutional networks. *Journal of Systems Architecture*, *95*(June 2018), 9–18. https://doi.org/10.1016/j.sysarc.2019.02.008

Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network

Training by Reducing Internal Covariate Shift. In F. Bach & D. Blei (Eds.), *Proceedings of the 32 nd International Conference on Machine Learning* (Vol. 37, pp. 448–456). http://proceedings.mlr.press/v37/ioffe15.html

Kim, S. Y., Ko, J. A., Kang, B. S., & Park, H. J. (2018). Prediction of key aroma development in coffees roasted to different degrees by colorimetric sensor array. *Food Chemistry*, *240*(July 2017), 808–816. https://doi.org/10.1016/j.foodchem.2017.07.139

Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations (ICLR 2015)*, 1–15. https://arxiv.org/abs/1412.6980

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, *25*, 1–9. https://doi.org/10.1016/B978-008046518-0.00119-7

Lathuiliere, S., Mesejo, P., Alameda-Pineda, X., & Horaud, R. (2020). A Comprehensive Analysis of Deep Regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(9), 2065–2081. https://doi.org/10.1109/TPAMI.2019.2910523

Le Cun, Y., Jackel, L. D., Boser, B., Denker, J. S., Graf, H. P., Guyon, I., Henderson, D., Howard, R. E., & Hubbard, W. (1989). Handwritten digit recognition: applications of neural network chips and automatic learning. *IEEE Communications Magazine*, *27*(11), 41–46. https://doi.org/10.1109/35.41400

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

Leme, D. S., da Silva, S. A., Barbosa, B. H. G., Borém, F. M., & Pereira, R. G. F. A. (2019). Recognition of coffee roasting degree using a computer vision system. *Computers and Electronics in Agriculture*, *156*(October 2018), 312–317. https://doi.org/10.1016/j.compag.2018.11.029

Li, H. X., & Si, H. (2017). Control for Intelligent Manufacturing: A Multiscale Challenge. *Engineering*, *3*(5), 608–615. https://doi.org/10.1016/J.ENG.2017.05.016

Liu, H., Lu, J., Feng, J., & Zhou, J. (2019). Ordinal Deep Learning for Facial Age Estimation. *IEEE Transactions on Circuits and Systems for Video Technology*, *29*(2), 486–501. https://doi.org/10.1109/TCSVT.2017.2782709

Liu, Y., Pu, H., & Sun, D. (2021). Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing

complex food matrices. *Trends in Food Science & Technology*, *113*(May), 193–204. https://doi.org/10.1016/j.tifs.2021.04.042

Madihah, K. Y. K., Zaibunnisa, A. H., Norashikin, S., Rozita, O., & Misnawi, J. (2012). Optimization of Roasting Conditions for High-Quality Robusta Coffee. *APCBEE Procedia*, *4*(4), 209–214. https://doi.org/10.1016/j.apcbee.2012.11.035

Meng, Q., Chen, W., Wang, Y., Ma, Z.-M., & Liu, T.-Y. (2019). Convergence analysis of distributed stochastic gradient descent with shuffling. *Neurocomputing*, *337*, 46–57. https://doi.org/10.1016/j.neucom.2019.01.037

Moreira, A. S. P., Nunes, F. M., Domingues, M. R., & Coimbra, M. A. (2012). Coffee melanoidins: structures, mechanisms of formation and potential health impacts. *Food & Function*, *3*(9), 903. https://doi.org/10.1039/c2fo30048f

Pfisterer, K. J., Amelard, R., Chung, A. G., & Wong, A. (2018). A new take on measuring relative nutritional density: The feasibility of using a deep neural network to assess commercially-prepared puréed food concentrations. *Journal of Food Engineering*, *223*, 220–235. https://doi.org/10.1016/j.jfoodeng.2017.10.016

Rafegas, I., & Vanrell, M. (2018). Color encoding in biologically-inspired convolutional neural networks. *Vision Research*, *151*(May), 7–17. https://doi.org/10.1016/j.visres.2018.03.010

Romani, S., Cevoli, C., Fabbri, A., Alessandrini, L., & Dalla Rosa, M. (2012). Evaluation of Coffee Roasting Degree by Using Electronic Nose and Artificial Neural Network for Off-line Quality Control. *Journal of Food Science*, *77*(9), C960–C965. https://doi.org/10.1111/j.1750-3841.2012.02851.x

Sacchetti, G., Di Mattia, C., Pittia, P., & Mastrocola, D. (2009). Effect of roasting degree, equivalent thermal effect and coffee type on the radical scavenging activity of coffee brews and their phenolic fraction. *Journal of Food Engineering*, *90*(1), 74–80. https://doi.org/10.1016/j.jfoodeng.2008.06.005

Santos, J. R., Viegas, O., Páscoa, R. N. M. J., Ferreira, I. M. P. L. V. O., Rangel, A. O. S. S., & Lopes, J. A. (2016). In-line monitoring of the coffee roasting process with near infrared spectroscopy: Measurement of sucrose and colour. *Food Chemistry*, *208*, 103–110. https://doi.org/10.1016/j.foodchem.2016.03.114

Schmittmann, O., & Lammers, P. (2017). A true-color sensor and suitable evaluation algorithm for plant recognition. *Sensors (Switzerland)*, *17*(8). https://doi.org/10.3390/s17081823

Shen, Y., Zhou, H., Li, J., Jian, F., & Jayas, D. S. (2018). Detection of stored-grain insects using deep learning. *Computers and Electronics in Agriculture*, *145*(June 2017), 319–325. https://doi.org/10.1016/j.compag.2017.11.039

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. In Y. Bengio & Y. LeCun (Eds.), *3rd International Conference on Learning Representations, ICLR 2015* (pp. 1–14). http://arxiv.org/abs/1409.1556

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *07-12-June*, 1–9. https://doi.org/10.1109/CVPR.2015.7298594

Tsai, S.-Y., Hwang, B.-F., Wang, S.-P., & Lin, C.-P. (2017). A Kinetics Study of Coffee Bean of Roasting and Storage Conditions. *Journal of Food Processing and Preservation*, *41*(4), e13040. https://doi.org/10.1111/jfpp.13040

Urwin, R., Kesa, H., & Joao, E. S. (2019). The rise of specialty coffee: An investigation into the consumers of specialty coffee in Gauteng. *African Journal of Hospitality, Tourism and Leisure*, *8*(5), 39–40.

USDA, U. S. D. of A. (2020). Coffee: World Markets and Trade. In *Coffee: World Markets and Trade*. http://apps.fas.usda.gov/psdonline/circulars/coffee.pdf

Vargas-Elías, G. A., Corrêa, P. C., Souza, N. R. de, Baptestini, F. M., & Melo, E. de C. (2016). Kinetics of mass loss of arabica coffee during roasting process. *Engenharia Agrícola*, *36*(2), 300–308. https://doi.org/10.1590/1809-4430-Eng.Agric.v36n2p300-308/2016

Virgen-Navarro, L., Herrera-López, E. J., Corona-González, R. I., Arriola-Guevara, E., & Guatemala-Morales, G. M. (2016). Neuro-fuzzy model based on digital images for the monitoring of coffee bean color during roasting in a spouted bed. *Expert Systems with Applications*, *54*, 162–169. https://doi.org/10.1016/j.eswa.2016.01.027

Wang, J., Ma, Y., Zhang, L., Gao, R. X., & Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems*, *48*, 144–156. https://doi.org/10.1016/j.jmsy.2018.01.003

Wang, X., & Lim, L.-T. (2014). A Kinetics and Modeling Study of Coffee Roasting Under Isothermal Conditions. *Food and Bioprocess Technology*, *7*(3), 621–632. https://doi.org/10.1007/s11947-013-1159-8

Wang, X., & Lim, L. T. (2017). Investigation of CO2 precursors in roasted coffee. *Food Chemistry*, *219*, 185–192. https://doi.org/10.1016/j.foodchem.2016.09.095

Wu, C. W. (2018). *ProdSumNet: reducing model parameters in deep neural networks via product-of-sums matrix decompositions*. http://arxiv.org/abs/1809.02209

Wu, H., & Zhao, J. (2018). Deep convolutional neural network model based chemical

process fault diagnosis. *Computers and Chemical Engineering*, *115*, 185–197. https://doi.org/10.1016/j.compchemeng.2018.04.009

Yong, Y. L., Tan, L. K., McLaughlin, R. A., Chee, K. H., & Liew, Y. M. (2017). Linear-regression convolutional neural network for fully automated coronary lumen segmentation in intravascular optical coherence tomography. *Journal of Biomedical Optics*, *22*(12), 1. https://doi.org/10.1117/1.JBO.22.12.126005

Yuan, X., Qi, S., Shardt, Y. A. W., Wang, Y., Yang, C., & Gui, W. (2020). Soft sensor model for dynamic processes based on multichannel convolutional neural network. *Chemometrics and Intelligent Laboratory Systems*, *203*(April), 104050. https://doi.org/10.1016/j.chemolab.2020.104050

Zhang, M. L., & Zhou, Z. H. (2014). A Review on Multi-Label Learning Algorithms. *IEEE Transactions on Knowledge and Data Engineering*, *26*(8), 1819–1837. https://doi.org/10.1109/TKDE.2013.39

Zhang, Q., Zhuo, L., Li, J., Zhang, J., Zhang, H., & Li, X. (2018). Vehicle color recognition using Multiple-Layer Feature Representations of lightweight convolutional neural network. *Signal Processing*, *147*, 146–153. https://doi.org/10.1016/j.sigpro.2018.01.021

Zhang, X., Li, X., Feng, Y., & Liu, Z. (2015). The use of ROC and AUC in the validation of objective image fusion evaluation metrics. *Signal Processing*, *115*, 38–48. https://doi.org/10.1016/j.sigpro.2015.03.007

Zhang, Yan, Lian, J., Fan, M., & Zheng, Y. (2018). Deep indicator for fine-grained classification of banana's ripening stages. *EURASIP Journal on Image and Video Processing*, *2018*(1), 46. https://doi.org/10.1186/s13640-018-0284-8

Zhang, Yiheng, Qiu, Z., Yao, T., Liu, D., & Mei, T. (2018). Fully Convolutional Adaptation Networks for Semantic Segmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 6810–6818. https://doi.org/10.1109/CVPR.2018.00712

Zhong, R. Y., Xu, X., Klotz, E., & Newman, S. T. (2017). Intelligent Manufacturing in the Context of Industry 4.0: A Review. *Engineering*, *3*(5), 616–630. https://doi.org/10.1016/J.ENG.2017.05.015

6. CONCLUSÕES GERAIS

Neste trabalho foi abordada a modelagem do processo de escurecimento não enzimático em amostras de café e pães, decorrente dos processos de torra e forneamento, respectivamente, utilizando técnicas de inteligência artificial, especialmente redes neurais convolucionais. Os resultados obtidos suportam as seguintes conclusões:

1. O uso de redes neurais convolucionais permite a construção de sistemas de visão computacional não dependentes de variáveis de processo de difícil mensuração tais como temperatura e umidade da superfície de uma amostra, o que representa uma grande vantagem frente aos modelos tradicionais.

2. As redes neurais convolucionais são capazes de extrair características de cores presentes em imagens, como aquelas vistas no processo de torra de café ou forneamento de pães, e utilizá-las para classificação dos estágios do processo.

3. Redes neurais convolucionais com reduzido número de camadas convolucionais apresentam melhor desempenho que modelos mais profundos ao lidarem com bancos de dados reduzidos e pequeno número de classes. Isso é particularmente útil para muitos casos observados na indústria de alimentos e na pesquisa, onde o tamanho do banco de dados costuma ser um problema e o número de classes geralmente é pequeno.

Embora tenha sido demonstrada a superioridade das RNC frente aos modelos tradicionais, algumas oportunidades para futuros trabalhos surgem exatamente da combinação entre eles. Em especial no desenvolvimento de sistemas mais complexos, como é o caso do gêmeo digital, o qual consiste numa cópia virtual fiel de um processo ou objeto físico. Tais sistemas representam o ápice da modelagem de um processo e permite a simulações em tempo real, com vista a predição de falhas, ajustes e otimização do mesmo. Porém, a modelagem fenomenológica é incapaz de capturar toda a complexidade presente no processamento de alimentos. Dessa forma, a combinação da modelagem fenomenológica com a modelagem com técnicas de inteligência artificial devem resultar em modelos robustos e representativos do processo.