

# Ciência de Dados na Indústria Vitivinícola: compreensão, pesquisa e maturação tecnológica

1<sup>st</sup> Priscila Cathlen Alves Sá  
Escola Superior de Agricultura Luis de Queiroz  
Universidade de São Paulo  
Piracicaba, Brasil  
orcid.org/0009-0005-3567-3170

2<sup>nd</sup> Ricardo Menezes Prates  
Colegiado de Engenharia Elétrica  
Universidade Federal do Vale do São Francisco  
Juazeiro-Ba, Brasil  
orcid.org/0000-0002-1580-9828

**Abstract**— This work describes the trajectory of an extension and research project in Applied Data Science within the industrial sector of viticulture. The process involves understanding a business problem, using bibliographic research as a tool to select the most suitable technology for an envisioned solution, as well as evaluating its technological maturity up to the pre-incubation stage of a startup. It explores how innovation environments drive the maturation process of new technologies and startups, promoting collaboration and competitiveness among productive sectors and educational and research institutions.

**Keywords**—*startup*, data science, business, industry.

**Resumo**— Este trabalho descreve a trajetória de um projeto de extensão e pesquisa em Ciência de Dados Aplicada ao setor industrial da vitivinicultura. O processo de compreender um problema de negócio, o uso da pesquisa bibliográfica como ferramenta para a escolha de tecnologia mais adequada para uma solução idealizada, assim como a avaliação de sua maturidade tecnológica até a pré-incubação de uma startup. Explora como ambientes de inovação impulsionam o processo de maturação de novas tecnologias e startups, promovendo colaboração e competitividade entre os setores produtivos e as instituições de ensino e pesquisa.

**Palavras-chave**— *startup*, ciência de dados, inovação, negócios, indústria.

## I. INTRODUÇÃO

### A. Ambientes de inovação para o desenvolvimento da ciência de dados: contexto de Petrolina -Pernambuco

Evoluir um projeto de pesquisa em um produto tecnológico, que seja, ao mesmo tempo, rentável e escalável, não é um processo que compete somente ao ambiente acadêmico. Um importante reflexo disso é a alta taxa de mortalidade de *startups* no período de seu ciclo de vida conhecido como “vale da morte” [1].

O ciclo de vida de uma *startup* é caracterizado por diferentes estágios, cada um com suas próprias necessidades de recursos e desafios [2]. O “vale da morte” é um termo utilizado para descrever a fase crítica na qual muitas *startups* falham por falta de recursos financeiros. Essa fase ocorre entre o estágio de pré-*startup* e o estágio de *startup*, quando a empresa precisa de capital para desenvolver e testar seu produto ou serviço, mas ainda não tem receita suficiente para se sustentar [1].

Muitos fatores estão associados ao sucesso ou fracasso de uma ideia inovadora de base tecnológica, entre outras: baixo nível de maturidade tecnológica do produto, deficiências na gestão de processos, falta de equipes e de fomento financeiro estão entre os principais indicadores [3]. E nesse cenário de incertezas e desafios os ambientes de inovação, como o *Vitis*

*Habitat* em Petrolina, são importantes atores de fomento da cultura, da economia e da troca de conhecimento sobre inovação e tecnologia entre os diversos *players* de um ecossistema de inovação [3, 4].

O “*Vitis Habitat* – Ambiente para Inovação do Arranjo Produtivo de Vinhos e Sucos” é um Lócus de Inovação credenciado em Pernambuco com fomento da Fundação de Amparo à Ciência e Tecnologia de Pernambuco (FACEPE), produto de alianças estratégicas de instituições públicas e privadas [5]. O lócus é um programa que visa a criação de um espaço físico e virtual onde os principais *stakeholders* da região de Petrolina: empreendedores, *startups*, empresas e investidores, possam se conectar, compartilhar recursos, conhecimentos, colaborar e desenvolver projetos, com o objetivo de impulsionar a inovação, o crescimento e sua competitividade no mercado.

Desde seu credenciamento, o *Vitis Habitat* tem importantes iniciativas, entre elas, o Programa de Extensão Tecnológica (PET), apoiado pela FACEPE, de Ciência de Dados Aplicada à Vitivinicultura de Precisão e Digital (CDA-VPD) [6]. Seu objetivo foi atuar junto à cadeia produtiva da indústria de vitivinicultura no Vale do São Francisco (VSF), na cidade de Petrolina no período entre 2022 e 2023 [6], desenvolvendo possíveis aplicações da ciência de dados a problemas reais de uma indústria. Essa parceria foi realizada por meio de documento específico de cooperação técnica entre os agentes de inovação e a indústria.

O projeto CDA-VPD tem seu percurso e resultados descritos nesse artigo para fins de evidenciar a trajetória de uma pesquisa de base tecnológica da ciência de dados aplicada à indústria, sua transformação em ideia de negócio, avaliação de maturidade com base na escala TRL e posterior aprovação em edital de fomento para pré-incubação de *startup*.

## II. DESENVOLVIMENTO: PROGRAMA DE EXTENSÃO E PESQUISA TECNOLÓGICA

### A. Contexto industrial: vitivinicultura na região do Sertão do São Francisco

A vitivinicultura, em especial o arranjo produtivo de vinho e sucos [7], é considerada como um setor estratégico para o estado de Pernambuco. O estado faz parte do VSF, hoje principal polo produtor de uvas de mesa do Brasil [8]. Esta região é o segundo maior produtor dessas bebidas no Brasil, se destacando por vinhos finos, espumantes e sucos com características especiais: a maior parte da produção se caracteriza por vinhos jovens, na grande maioria não reserva, para consumo em poucos meses, com mercado bem definido [9]. Esta condição favorável e particular se deve, principalmente, às características edafoclimáticas únicas, a

exemplo dos tipos de solo, radiação solar e precipitação pluvial [9]. Deste modo, encontra-se um *terroir* exclusivo, sendo o único local do mundo com produção de vinhos em condição de semiárido [10].

As indústrias de vinhos e sucos visitadas no Sertão do São Francisco - empresas parceiras por meio de cooperação técnica com o PET CDA-VPD - possuem diferentes escalas de amadurecimento e porte: desde pequenas produções industriais e orgânicas, até grandes corporações e grupos internacionais. Suas atividades são caracterizadas por ser agroindustrial, pois ao longo da sua cadeia produtiva a matéria prima, as uvas, são produzidas em campo e transportadas para uma planta industrial para o processamento e transformação em produtos. Trata-se de um processo de produção desafiador, pois o produto depende da qualidade da matéria prima, as uvas. Por conseguinte, o monitoramento da qualidade ao longo do processo é imprescindível para aumentar a produtividade, como também para contribuir na tomada de decisões mais assertivas no processo industrial [11].

### B. Ciência de dados aplicada a vitivinicultura de precisão e digital (CDA-VPD)

O programa teve duas etapas: a primeira foi um Curso de Capacitação em Ciência de Dados aplicada à agroindústria, ofertado de maneira híbrida com carga horária total de quarenta horas e com quarenta alunos selecionados em edital específico. O perfil da turma variou entre pessoas com conhecimentos avançados a alunos que ainda não possuíam conhecimentos básicos em inteligência computacional.

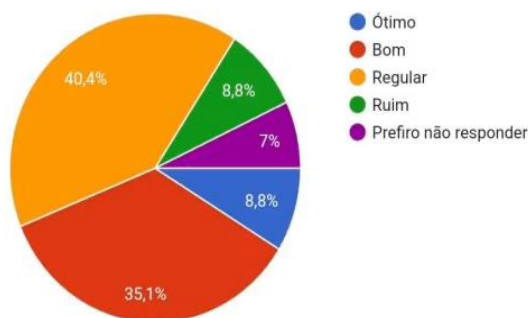


Fig. 1. Perfil da turma.

O curso tinha por objetivos principais desenvolver competências e habilidades dos alunos na linguagem de programação *Python* para aprendizado de máquina, modelos computacionais inteligentes, investigação e implementação de potenciais sistemas inteligentes.

A segunda etapa do programa consistiu na seleção de dez alunos para uma bolsa de pesquisa e imersão em uma vinícola industrial para conhecer os processos de trabalho e gestão com objetivo de idealizar aplicações de modelos computacionais inteligentes.

### C. Compreensão do problema de negócio - Imersões

Quando se pretende que o produto/conhecimento de uma pesquisa científica resulte em um produto de mercado, parte de seu desenvolvimento precisa considerar um problema real de negócio [4]. Uma boa solução sem demanda de aplicação pode falhar em se sustentar enquanto solução industrial.

Partindo dessa premissa, antes da escolha acadêmica de qual técnica em Ciência de Dados a ser utilizada no PET, realizou-se uma pesquisa técnica e imersiva para levantamento de informações relevantes sobre o funcionamento de negócio da indústria parceira do PET, durante o mês de fevereiro de 2023. As visitas forneceram importantes *insights* sobre a “dor do negócio” [3,4] e partiu-se daí para a escolha da técnica e solução que melhor se adequaria àquele contexto a partir de uma pesquisa bibliográfica de uso de modelos inteligentes na vitivinicultura.

### D. Um problema real

Bons indicadores de uma aplicação potencial de tecnologia podem ser conseguidos com a identificação e mapeamento de atividades relacionadas a processos de trabalho industrial com alto potencial de serem automatizáveis: trabalhos manuais ou que produzem grandes volumes de informações, de fontes diversas e organizadas em sistemas diferentes de planilhas e *softwares*, que geralmente não se comunicam, indicam para um pesquisador locais onde é possível contribuir com soluções e processos que tenham impacto e adesão a realidade de uma indústria[4].

Foram identificados e mapeados por entrevistas e visitas locais de trabalho que a indústria parceira tinha grandes preocupações e custos com processos manuais ou semiautomáticos nas atividades de avaliação de qualidade de uvas produzidas e a consequente qualidade de vinhos devida à alta correlação entre esses fatores. O projeto atentou-se a duas atividades realizadas no ambiente industrial agrícola: manipulação de qualidade dos frutos no *packing house* e existência de um banco de dados de informações do laboratório da indústria sobre características químicas das uvas produzidas, como: acidez fixa, açúcar residual, cloreto, dióxido de enxofre livre, densidade e pH.

Identificou-se alguns potenciais problemas de negócio que pudessem ter uma solução desenvolvida por meio da ciência de dados, entre outros: o primeiro problema era que apesar de ter uma extensa base de dados sobre características químicas das uvas, a indústria não possuía qualquer programa interno para tratamento, manipulação e compreensão de dados para gerar valor ao negócio. O segundo problema era a alta dependência de mão de obra humana nas atividades do *packing house*, principalmente, as relacionadas a identificação de qualidade físicas das frutas.

Identificado esse cenário e “dor de negócio”, construiu-se um documento técnico com duas possíveis soluções baseadas em IA. O objetivo do documento técnico era ser um informativo para empresas sobre as possibilidades de aplicação dessas técnicas para agregar aos processos industriais maior produtividade, diminuir erros de avaliação e auxiliar na tomada de decisões baseadas em dados. Essas propostas foram elaboradas com a intenção de oferecer às empresas uma visão clara de como a IA pode ser aproveitada para otimizar seus processos, aumentar a eficiência e a precisão das avaliações, além de fornecer um suporte valioso para a tomada de decisões estratégicas fundamentadas em informações verificáveis.

Para desenvolvimento deste trabalho, exploraremos em detalhes pesquisas científicas para a escolha da tecnologia que melhor atenda a solução idealizada a ser aplicada no processo de beneficiamento de frutos, também conhecido como *packing houses*.



Figura 2. Etapas de construção do relatório técnico. Fonte: própria autoria.

### E. Packing house: identificando um potencial produto de negócio

A produção industrial de frutos é composta por muitas etapas, que englobam atividades do campo até a planta industrial de uma fazenda produtora [12]. A finalidade desses negócios, em suma, é cultivar boas safras e embarcar os frutos ou produtos para seu mercado consumidor. Para viabilizar a venda de toda essa safra e minimizar os danos ocorridos após a colheita, é imprescindível para essas indústrias processar os frutos, o que normalmente é realizado em instalações de empacotamento conhecidas como *packing houses*. Dentro dessas edificações são executadas tarefas de recebimento dos frutos, higienização, categorização, embalagem e transformação em pellets [12,13].



Fig. 3. Packing houses. Fonte: LUZ, 2016.

Também conhecidas como usinas de beneficiamento de frutos, esses espaços são caracterizados por suas dimensões expansivas e instalações projetadas para otimizar o fluxo de materiais e a eficiência dos funcionários. O processo de beneficiamento das frutas em *packing houses* é dividido em várias etapas distintas, as quais compreendem:

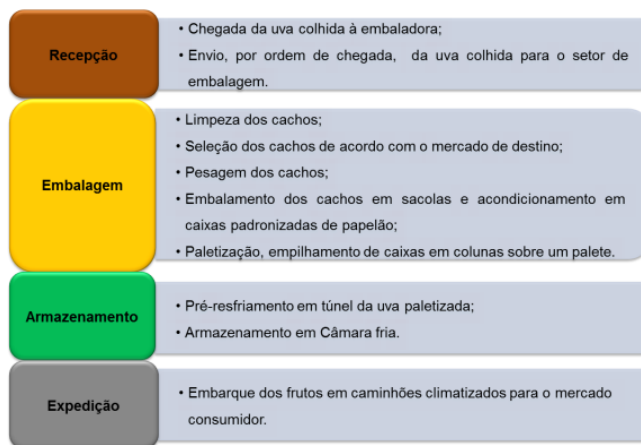


Fig. 4. Etapas de beneficiamento do packing houses. Fonte: LUZ, 2016.

Esse processo tem como principal objetivo garantir que as frutas atendam em qualidade física e química o nível de exigência de seu mercado consumidor, geralmente mercados externos. Segundo a Associação Brasileira dos Produtores e Exportadores de Frutas e Derivados (ABRAFRUTAS), em 2021, Petrolina exportou mais de 170 mil toneladas de manga e 50 mil toneladas de uva para mercados dos Estados Unidos, Reino Unido e Espanha [14].

O grupo de pesquisa identificou na etapa de embalagem, especificamente na atividade de seleção de frutos, a oportunidade de negócio: aprimorar as mesas de decisão do *packing houses* com visão computacional na etapa de seleção de frutos.

### III. METODOLOGIA: PESQUISA BIBLIOGRÁFICA

Após a identificação do problema, demos início ao processo de pesquisa por estudos científicos. Estas pesquisas podem ajudar a compor sistemas inteligentes comerciais de interesse da indústria 4.0. Nosso objetivo era verificar a viabilidade tecnológica da visão computacional na identificação e segmentação de imagens de uvas, a fim de resolver o nosso problema de negócio.

A seguir, apresentamos um resumo de metodologias e resultados dos artigos de interesse selecionados e assim auxiliar na escolha da solução a ser desenvolvida.

Os critérios de seleção foram baseados em palavras-chave relevantes, incluindo "inovação", "uvas" e "modelos inteligentes". Foi dada prioridade a artigos que apresentassem técnicas diversas e inovadoras, no período de 2015 a 2023.

#### A. "Non-invasive setup for grape maturation classification using deep learning"

Este estudo propõe o uso de redes neurais convolucionais (CNNs) para classificar a maturação de uvas de forma não invasiva [15]. O objetivo do estudo foi determinar o momento ideal de colheita das uvas *Syrah* e *Cabernet Sauvignon* usando técnicas de processamento de imagens e aprendizado de máquina.

O estudo utiliza uma abordagem baseada em imagens adquiridas por câmera óptica *Canon EOS REBEL T5i*. As uvas foram fotografadas e processadas por algoritmos de aprendizado profundo: VGG-19 e CNN. Essas redes neurais são capazes de aprender padrões e características nas imagens das uvas, permitindo a classificação da maturação. Para validar a eficácia do método proposto, o estudo realizou experimentos com amostras de uvas.

As uvas foram maceradas para facilitar a extração dos compostos desejados para determinar a concentração de antocianinas. As amostras foram armazenadas em recipientes protegidos da luz e refrigerados. Após 24 horas, o sobrenadante foi coletado e analisado em um espectrofotômetro modelo *Spectronic BioMate 5 UV-Vis* em um comprimento de onda de 374 nm. Os resultados foram expressos em miligramas por 100g de uva.

O desempenho da VGG-19 e da CNN foram comparados em diferentes cenários. Para o conjunto de dados 'Syrah' em um cenário de três classes, a CNN apresentou uma precisão média de teste de 93,41%. Já para o conjunto de dados 'Cabernet Sauvignon' em um cenário de quatro classes, a CNN proposta obteve uma precisão média de teste de 72,66%. Para o cenário de oito classes, a VGG-19 obteve uma acurácia de validação de 80,69% e uma acurácia de teste de 76,23%. Já a CNN proposta alcançou uma acurácia de validação de 64,52% e uma acurácia de teste de 65,30%. Para o cenário de quatro classes, a VGG-19 obteve uma acurácia de validação de 81,63% e uma acurácia de teste de 80,97%. A CNN proposta alcançou uma acurácia de validação de 79,28% e uma acurácia de teste de 72,66%. Portanto, com base nas informações fornecidas, a CNN apresentou resultados de acurácia variados em comparação com a VGG-19, dependendo do cenário e da cultivar considerados.

Concluiu-se que a utilização de CNNs e técnicas de processamento de imagens pode ser uma alternativa viável e eficaz para a classificação da maturação de uvas, oferecendo benefícios tanto em termos de precisão quanto de economia de recursos.

#### B. "Predicting Grape Sugar Content under Quality Attributes Using Normalized Difference Vegetation Index Data and Automated Machine Learning"

O estudo destaca as descobertas promissoras na aplicação de técnicas de aprendizado de máquina automatizado (AutoML) de código aberto e dados de índice de vegetação de diferença normalizada (NDVI) [16], apresentando um estudo sobre a previsão do teor de açúcar das uvas com base em atributos de qualidade, utilizando dados NDVI e aprendizado de máquina automatizado.

Os autores realizaram a coleta de dados, o pré-processamento dos dados e o desenvolvimento de metodologia para o estudo: um método que combina técnicas de AutoML de código aberto e dados NDVI para prever a qualidade das uvas. Os dados NDVI foram coletados em diferentes estágios de crescimento das vinhas, utilizando métodos não destrutivos, como sensores de refletância de dossel montados em veículos próximos, imagens de UAV e imagens de sensoriamento remoto *Sentinel-2*. Foram exploradas diferentes configurações, incluindo diferentes combinações de sensores e dados coletados ao longo de duas temporadas de crescimento, 2019 e 2020. Além disso, foi feita uma comparação entre os métodos de aprendizado de máquina ajustados manualmente e o AutoML, utilizando Máquinas de Vetores de Suporte (SVMs) e Determinação Automática de Relevância (ARD).

Os resultados do estudo revelaram que tanto o método manual de ajuste de aprendizado de máquina quanto o AutoML foram capazes de obter resultados semelhantes na previsão da qualidade das uvas. No entanto, o AutoML apresentou uma pequena vantagem em relação ao método manual para os anos de 2019 e 2020. Além disso, ao combinar

múltiplos sensores e estágios de crescimento ao longo do tempo, houve uma melhoria ainda maior no coeficiente de determinação  $R^2$ . Os sensores aéreos, juntamente com os estágios de *Véraison* e *Floração*, demonstraram ter os maiores valores médios de  $R^2$ . Diferentes algoritmos de regressão testados com o AutoML também mostraram resultados superiores em comparação com o método manualmente ajustado. Esses resultados indicam que o uso do AutoML pode aumentar a capacidade preditiva e fornecer uma solução mais confiável e sustentável para a previsão da qualidade das uvas.

É importante considerar alguns aspectos relevantes com base nas informações fornecidas na discussão do estudo. Primeiramente, é necessário discutir a eficácia do AutoML em comparação com os métodos convencionais de ajuste manual de aprendizado de máquina. Além disso, é fundamental analisar o impacto da combinação de diferentes sensores e estágios de crescimento na precisão das previsões. A validação cruzada e a repetição dos experimentos também devem ser abordadas para avaliar a generalização e a robustez dos modelos. Além disso, é importante considerar as implicações da restrição de acesso aos dados e como isso pode afetar a replicabilidade da pesquisa. Por fim, é necessário identificar quaisquer limitações do estudo e propor possíveis direções futuras de pesquisa para abordar essas limitações e expandir o conhecimento nessa área.

#### C. "A Non-Invasive Method Based on Computer Vision for Grapevine Cluster Compactness Assessment Using a Mobile Sensing Platform under Field Conditions"

Este artigo descreve um método inovador para avaliar a compactação de cachos de videira usando uma plataforma de sensoriamento remoto móvel [17]. A plataforma, desenvolvida na Universidade de *La Rioja*, é montada em um veículo todo-terreno modificado e possui uma câmera RGB (vermelho, verde, azul), um sistema de posicionamento global (GPS) e um sistema de iluminação artificial. As imagens das copas das videiras e das zonas de frutificação são capturadas durante a noite, garantindo uma iluminação homogênea e separando os cachos em avaliação dos cachos da fileira oposta.

As imagens adquiridas são processadas por meio de algoritmos de visão computacional e aprendizado de máquina. Esses algoritmos analisam a porosidade da copa da videira, a exposição dos cachos e das folhas, e a compactação dos cachos. O artigo não menciona explicitamente qual algoritmo de visão computacional foi utilizado. No entanto, é mencionado que o método combina visão computacional no processamento de imagem, como detecção de bordas, segmentação de objetos e extração de características, juntamente com algoritmos de aprendizado de máquina, como regressão ou classificação, para realizar a análise das imagens e a avaliação da compactação dos cachos.

O método foi validado usando um conjunto de dados de 100 cachos de uva de diferentes variedades e vinhedos.

Os resultados obtidos demonstram a eficácia desse método na avaliação da compactação dos cachos de videira em termos de sensibilidade, especificidade, pontuação F1, AUC e IoU. Especificamente, as classes mais relevantes, como "*rachis*" e "*grape*" apresentaram valores de especificidade semelhantes e igualmente bons em todos os conjuntos de dados, enquanto os resultados foram mais variáveis para as métricas restantes.

O estudo menciona algumas restrições que podem impactar a estimativa da compactação dos cachos de uva usando a metodologia proposta. A presença de folhas ocultando parte do cacho pode levar a uma estimativa imprecisa de compactação. Além disso, quando os cachos estão sobrepostos, pode ocorrer a identificação errônea de múltiplos cachos como um único, prejudicando a estimativa individual de compactação. Essas limitações destacam a importância de considerar as condições do campo ao aplicar o método e interpretar os resultados obtidos.

#### D. “Grape Berry Detection and Size Measurement Based on Edge Image Processing and Geometric Morphology”

O objetivo deste estudo é criar um método para detectar e medir o tamanho das bagas de uva usando técnicas de processamento de imagem [18]. O estudo busca melhorar a precisão na contagem e caracterização das bagas de uva.

O estudo utiliza técnicas de processamento de imagem para detectar as bagas de uva e medir seu tamanho. O método proposto é baseado em ajuste de mínimos quadrados e utiliza informações parciais das bagas para inferir a forma completa, contando e caracterizando as bagas de uva. As imagens utilizadas neste estudo foram coletadas no Parque Ecológico de Guangdong Luonan e consistem em diferentes tipos de uvas, incluindo Niagara, uvas pretas e uvas vermelhas. Essas imagens foram capturadas em condições climáticas favoráveis, e um sistema de aquisição de imagens foi utilizado, composto por uma câmera e um tripé. A câmera possui uma resolução máxima de  $1280 \times 480$  pixels. Essas imagens foram usadas para testar e avaliar o método proposto de detecção e medição do tamanho das bagas de uva.

Os resultados experimentais mostram que o método proposto é capaz de detectar com precisão as bagas de uva mesmo quando estão sobrepostas. A precisão média na medição do tamanho das bagas é de 2,30 mm, com um erro máximo de 5,62 mm. Apesar de fatores como ruído, oclusão e curvatura da interseção afetarem os resultados, a precisão na detecção das bagas de uva é melhorada. O estudo conclui que o método proposto pode ser usado para estimar com precisão o peso, o rendimento e a qualidade das uvas individuais por meio da tecnologia de percepção visual em um ambiente natural.

O método proposto para detectar e medir o tamanho das bagas de uva apresenta algumas restrições. Ele foi desenvolvido principalmente para bagas de uva que possuem uma forma arredondada ou quase circular, e sua eficácia pode ser comprometida ao lidar com bagas que possuem formas diferentes ou alongadas. Embora o método tenha demonstrado uma precisão média na medição do tamanho das bagas, é importante levar em consideração essas limitações ao aplicá-lo em diferentes situações.

#### E. “Grape Leaf Disease Identification using Machine Learning Techniques”

Este trabalho descreve um sistema automatizado que foi desenvolvido com o objetivo de identificar e classificar doenças em folhas de uva [19]. Para alcançar esse objetivo, foram utilizadas técnicas avançadas de processamento de imagem e aprendizado de máquina.

O sistema utiliza um algoritmo chamado *GrabCut* para separar a região de interesse das folhas do restante da imagem. Em seguida, são extraídas características globais das imagens para realizar a classificação. Diferentes algoritmos de

aprendizado de máquina, como SVM, *Adaboost* e *Random Forest*, são aplicados para classificar as folhas segmentadas.

Os resultados obtidos mostram que o sistema alcançou uma precisão de classificação de até 91% utilizando o algoritmo SVM. Além disso, o método de aprendizado semissupervisionado obteve uma precisão média de 86,57%. O trabalho também apresenta imagens ilustrativas das diferentes doenças encontradas nas folhas de uva. Mostrando-se eficaz na detecção e classificação de doenças em folhas de uva.

Importante considerar as possíveis limitações: variações nas condições de iluminação das imagens, a possibilidade de não abranger todas as doenças das folhas de uva, desafios na aplicação do sistema em outras culturas e a dependência da disponibilidade de dados de treinamento adequados.

#### F. “Assessment of grape cluster yield components based on 3D descriptors using stereo vision”

Este estudo propõe uma abordagem de visão computacional tridimensional para estimar os componentes de rendimento de cachos de uva [20]. A tecnologia de visão estereoscópica e descritores 3D são utilizados para obter uma avaliação não invasiva e precisa da qualidade das uvas.

Foram utilizadas 100 amostras de cachos de uva de dez variedades de uva vermelha (*Vitis vinifera* L.), incluindo Miguel de Arco, *Monastrell*, *Bobal*, *Cabernet Franc*, *Planta Mula*, *Ruby\_Seedless*, *Mouraton*, *Naparo*, *Moravia Agria* e *Moscatel Morisco* [21].

O estudo utiliza um sistema de duas câmeras estereoscópicas para capturar imagens das uvas, que são usadas para reconstruir parcialmente um modelo 3D do cacho. As imagens são obtidas utilizando um programa chamado *PGR FlyCapture v.1.22* (Point Grey Research Inc, Richmond, BC, Canadá). Os descritores 3D extraídos são então utilizados para estimar a qualidade do cacho e outros componentes de rendimento.

Embora o estudo tenha utilizado amostras de variedades de uva vermelha, é importante considerar possíveis limitações, como o tamanho da amostra e a variação nas condições de iluminação. No entanto, essa abordagem mostra potencial para ser aplicada em outras culturas além das uvas, oferecendo uma alternativa eficiente e não invasiva para a estimativa da qualidade das videiras.

#### G. “A Multicultural Approach for Grape Bunch Weight Estimation Using Image Analysis”

O objetivo deste estudo foi aprimorar a precisão da estimativa de peso de cachos de uva por meio de uma abordagem multicultural baseada em análise de imagem [22]. Para isso, foram utilizadas cultivares autóctones brancas portuguesas, como Encruzado e Arinto, enxertadas em porta-enxertos específicos, além da cultivar *Syrah*.

As imagens dos cachos de uva foram coletadas em um ambiente de laboratório, utilizando uma câmera comercial *Nikon D5200* montada em um tripé. Um fundo azul ou branco foi utilizado para melhorar a segmentação dos cachos com base na cor. Os cachos foram pesados, fotografados e o número total de bagas por cacho foi contado. A análise de imagem foi utilizada para extrair características dos cachos de uva, como tamanho, forma e cor. A abordagem multicultural foi aplicada para desenvolver modelos de regressão que



relacionam as características dos cachos de uva com seu peso real.

Os resultados indicaram que a abordagem multicultivar proposta foi capaz de estimar com precisão o peso dos cachos de uva. Os modelos de regressão desenvolvidos apresentaram uma boa correlação entre as características dos cachos e seu peso real, evidenciado pelas medidas de  $R^2$ , RMSPE (*Root Mean Square Percentage Error*), Bias e EF (*Modeling Efficiency*). As cultivares autóctones brancas portuguesas, como Encruzado e Arinto, enxertadas em porta-enxertos específicos, além da cultivar *Syrah*, foram utilizadas no estudo. As imagens foram obtidas em um ambiente de laboratório, garantindo a consistência e a qualidade dos dados coletados.

Embora o estudo não tenha mencionado explicitamente suas limitações, é possível identificar algumas possíveis restrições com base nas informações fornecidas. Essas limitações incluem a falta de generalização dos resultados para outras variedades de uva, a coleta das imagens em um ambiente controlado, o tamanho da amostra utilizado e as possíveis limitações na precisão da extração das características dos cachos de uva por meio da análise de imagem.

#### H. “Grape Cluster Real-Time Detection in Complex Natural Scenes Based on YOLOv5s Deep Learning Network”

Este documento apresenta um estudo sobre a detecção em tempo real de aglomerados de uvas em ambientes naturais complexos [23]. O objetivo é desenvolver um método preciso e eficiente para identificar aglomerados de uvas usando o algoritmo YOLOv5s.

Os métodos utilizados neste estudo incluem o uso do YOLOv5s como rede de detecção e um conjunto de dados chamado *Grape-internet* para treinamento e validação. O desempenho dos modelos foi avaliado usando cinco indicadores: precisão, recall, mAP (*mean average precision*), F1 e velocidade de detecção. No conjunto de dados *Grape-internet* utilizado neste estudo, todas as imagens de uva foram coletadas da internet. A coleta abrangeu variedades como *Kyoho*, *Summer Black*, *Cabernet Sauvignon*, *Midnight Beauty*, *Manicure Finger*, *Fujiminori* e *Syrah*. Um total de 787 imagens de uvas. Posteriormente, essas imagens foram aleatoriamente recortadas em diferentes resoluções, variando de  $514 \times 460$  pixels a  $4160 \times 3120$  pixels.

Os principais resultados obtidos foram os seguintes: a precisão, recall, mAP e F1 do método proposto foram de 99,40%, 99,40%, 99,40% e 99,40%, respectivamente. Além disso, a velocidade de detecção alcançou 344,83 fps, demonstrando a capacidade do YOLOv5s de atender aos requisitos de alta precisão, alta velocidade e desempenho leve para a detecção de aglomerados de uvas.

O estudo comparou o desempenho de quatro redes de detecção de objetos de ponta a ponta diferentes para a detecção de aglomerados de uvas com o desempenho da YOLOv5s. As redes comparadas são YOLOv3, YOLOv4, Faster R-CNN e RetinaNet. O desempenho é avaliado em termos de precisão, recall, mAP e F1. o modelo YOLOv4 obteve o melhor desempenho geral, seguido pelo modelo YOLOv3. O modelo Faster R-CNN obteve o pior desempenho geral. A YOLOv5s obteve precisão de 1,84% superior ScaledYOLOv4-CSP.

Ao comparar com os outros três algoritmos de rede, observou-se que a precisão, o recall, o mAP e os valores de F1 da rede YOLOv5s não alcançaram os mais elevados, porém foram bastante similares. Essa similaridade ocorreu devido à redução da profundidade e largura da rede, que resultou em uma diminuição na precisão em troca de ganhos em velocidade e redução no tamanho da rede.

## IV. RESULTADOS

A partir da pesquisa bibliográfica foi possível escolher o modelo teórico que melhor atendesse a solução do problema de negócio estabelecido no projeto. A seguir as etapas que envolveram a escolha do modelo teórico, a avaliação da maturidade tecnológica da solução e a aplicação dos resultados para edital de pré-incubação da *startup*.

### A. Tabela

a) Quadro geral de metodologias e resultados da pesquisa bibliográfica:

TABELA I. QUADRO GERAL

Síntese da pesquisa bibliográfica		
Título Artigo <sup>a</sup>	Métodos	Resultados
N-ISG (2020)	CNNs(VGG-19 x CNN)	A CNN proposta alcançou uma acurácia de validação de 79,28% e uma acurácia de teste de 72,66%. Portanto, com base nas informações fornecidas, a CNN apresentou resultados de acurácia variados em comparação com a VGG-19, dependendo do cenário e da cultivar considerados.
PGSC (2022)	AutoML + NDVI	Os resultados demonstraram que tanto os algoritmos de aprendizado de máquina ajustados manualmente quanto a abordagem AutoML foram eficazes na previsão da qualidade das uvas. A abordagem AutoML, em particular, mostrou uma melhora marginal no coeficiente de determinação e uma redução no erro quadrático médio.
AN-IMB (2019)	R.G.B. + COMPUTER VISION	O sistema alcançou um coeficiente de determinação de 0,68 e um erro médio quadrático de 0,96 no conjunto de testes.
GBDS (2021)	COMPUTER VISION	Os resultados mostraram uma alta precisão, com um erro médio de 2,30 mm ao comparar o tamanho medido com o tamanho real das bagas.
GLDI (2019)	Grabcut+ SVM+RNA	Acurácia geral de 93,035% na classificação das doenças em 1135 imagens de teste.
AGCY (2015)	Visão estereoscópica e descritores 3D	Descritores 3D na estimativa precisa dos componentes de rendimento dos cachos de uva.
AMAG (2022)	Abordagem multicultivar + modelos de regressão	Boa correlação entre as características das imagens e o peso real dos cachos.
GCR-T (2022)	algoritmo YOLOv5s	Capacidade do YOLOv5s de atender aos requisitos de alta precisão, alta velocidade e desempenho leve para a detecção de aglomerados de uvas.

a. Sigla faz referência às primeiras 4 palavras do título do artigo seguido do ano de publicação.

Fig. 5. Síntese da pesquisa bibliográfica.

### B. Escolha do melhor modelo para o problema de negócio

Com base nas informações extraída dos diferentes estudos relacionados ao uso de modelos inteligentes para identificação de padrões em uvas, incluindo suas metodologias e resultados, e considerando também as demandas específicas de *packing houses*, o modelo considerado mais adequado para criar uma solução em visão computacional para essa aplicação seria o

GCR-T, que utiliza o algoritmo *YOLOv5s* para a detecção de aglomerados de uvas. O *YOLOv5s* é conhecido por sua capacidade de alta precisão, alta velocidade e desempenho leve, o que o torna ideal para o ambiente de um *packing house*.

A justificativa para essa escolha é baseada em diversos fatores. Primeiramente, o *YOLOv5s* demonstrou resultados promissores em estudos anteriores, inclusive em comparação com outros modelos, mostrando uma alta precisão na detecção de aglomerados de uvas. Além disso, sua velocidade de processamento é importante em um contexto de produção em larga escala, onde a eficiência é essencial.

### C. Transformação do projeto em resultados: produto para edital da Incubadora Tecnológica da UNIVASF

Conhecido um problema de negócio e identificado uma tecnologia capaz de atender essa demanda, a próxima etapa foi conseguir avaliar o nível de maturidade tecnológica dessa solução. O TRL, ou Nível de Prontidão Tecnológica, é um sistema de medição que avalia o grau de maturidade de uma tecnologia específica [24, 25]. Ele fornece uma estrutura para comparar de forma consistente o nível de maturidade entre diferentes tipos de tecnologia. Os TRLs são divididos em nove níveis, desde a observação e relato dos princípios básicos até a comprovação da tecnologia por meio de operações de missão bem-sucedidas. Estes níveis ajudam a determinar o estágio de desenvolvimento de uma tecnologia e a identificar os próximos passos necessários para sua implementação.

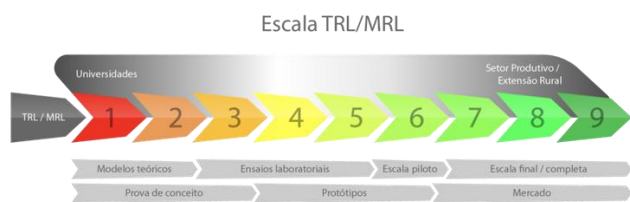


Fig. 6. Escala TRL. Fonte: NIT-UNIVASF, 2021.

Apesar do projeto ter tido a oportunidade de estar imerso em um setor produtivo real, em termos de maturidade, a tecnologia ainda é imatura, podendo ser localizada na escala no campo universidade no TRL, realizando a primeira etapa da prova de conceito: os modelos teóricos.

A partir desse ponto, os ecossistemas de inovação são também importantes atores de apoio para as recentes empresas avançarem nessa escala, pois muitos projetos esbarram nas limitações orçamentárias e técnicas.

O presente projeto foi aplicado ao Programa de Jornada Empreendedora da Incubadora Tecnológica da Univasf [26] e obteve aprovação para processo de pré-incubação no edital nº 09/2023 [27] como empresa intitulada “MAPA: agricultura de precisão”.

A Incubadora de Empresas de Base Tecnológica da Universidade Federal do Vale do São Francisco (Univasf), conhecida como INTECVASF, é a entidade encarregada de fomentar o desenvolvimento de *startups* de base tecnológica, promovendo a cultura empreendedora [28]. Além disso, tem como objetivo acolher empresas que possuam potencial para criar produtos, processos ou serviços inovadores, com alto valor agregado e viabilidade de aplicação no mercado.

## II CONCLUSÃO

A transformação de pesquisas tecnológicas em produtos tecnologicamente maduros e rentáveis é um processo

desafiador, mas essencial para impulsionar o desenvolvimento da ciência de dados. Os pesquisadores enfrentam obstáculos como a baixa barreira de conhecimento, a maturidade tecnológica do produto, a gestão de processos, as equipes e o fomento financeiro. No entanto, os ambientes de inovação desempenham um papel crucial nesse processo. Eles promovem a cultura da inovação, a troca de conhecimento e a colaboração entre os diversos atores do ecossistema. Os ambientes de inovação, como o *Vitis Habitat* em Petrolina, fornecem um espaço físico e virtual onde empreendedores, *startups*, empresas e investidores podem se conectar, compartilhar recursos, conhecimentos e colaborar no desenvolvimento de projetos.

A pesquisa bibliográfica desempenha um papel fundamental na escolha do modelo teórico adequado para solucionar os problemas de negócio. Além disso, a avaliação da maturidade tecnológica baseada na escala TRL ajuda a identificar as etapas necessárias para avançar no desenvolvimento do produto.

Portanto, é essencial aproveitar os ambientes de inovação para transformar pesquisas em produtos rentáveis. Ao superar os desafios de ideação e implementação de uma solução, aproveitar os recursos disponíveis, os pesquisadores podem impulsionar o desenvolvimento da ciência de dados, contribuindo para o crescimento econômico e a competitividade da região. A colaboração entre academia, indústria e investidores é fundamental para o sucesso desse processo.

## REFERÊNCIAS

- [1] J. Paschen, "Choose wisely: crowdfunding through the stages of the startup life cycle," *Business Horizons*, vol. 60, pp. 179-188, 2017.
- [2] J. A. Tessaro, R. Harms, and H. Schiele, "How startups become attractive to suppliers and achieve preferred customer status: Factors influencing the positioning of young firms," *Industrial Marketing Management*, vol. 113, pp. 100-115, Jun. 2023.
- [3] M. F. Neves, A. W. Gray, et al., "Ferramentas para o futuro do agro: estratégias para posicionar o Brasil como fornecedor mundial sustentável de alimentos, bioenergia e outros agroprodutos," São Paulo: Editora Gente, 2021, pp. 1-320.
- [4] SEBRAE/MG. "Startups do Agronegócio: método de criação, planejamento e gestão de startups do agronegócio". Belo Horizonte: SEBRAE/MG, 2020, 168p.
- [5] SECTI - Secretaria de Ciência, Tecnologia e Inovação do Estado de Pernambuco. "Edital FACEPE N. 02/2022 Credenciamento de Ambientes para Inovação em Pernambuco - Lócus de Inovação 2022", 2022.
- [6] SECTI - Secretaria de Ciência, Tecnologia e Inovação do Estado de Pernambuco. "Programa De Extensão Tecnológica Aprovados n. Edital Facepe Nº 22/2022", 2022.
- [7] L. E. M. Távora and A. V. de A. Cavalcanti, "Arranjo produtivo de viticultura, vinhos e derivados em Pernambuco: plano de melhoria da competitividade," Recife: SECTI/PE, Secretaria de Ciência, Tecnologia e Inovação do Estado de Pernambuco, 2017.
- [8] M. F. Moura, J. L. Hernandez, and M. Pedro Júnior, "Uvas de interesse econômico para vinificação e consumo in natura," *Visão agrícola*, vol. 14, pp. 8-14, junho 2021.
- [9] M. R. V. da Silva, "A Vitivinicultura e o Enoturismo no Vale do Submédio São Francisco," *Dissertação de mestrado*, Universidade Federal Rural de Pernambuco, Recife, Novembro 2018.
- [10] Regiões vinícolas brasileiras avançam na obtenção de Indicação de Procedência e Denominação de Origem, o terroir dos vinhos," *Visão agrícola*, vol. 14, pp. 57-60, junho 2021.
- [11] A. F. Nogueira Junior, L. Amorim, e M. B. Spósito, "Videiras requerem monitoramento e combater às doenças de início e fim de ciclo," *Visão Agrícola*, vol. 14, nº 3, pp. 71-75, jun. 2021.

- [12] S. N. Luz, "Caracterização de packing houses de uvas na região de Petrolina/Juazeiro e proposição de uma lista de verificação para o desenvolvimento de projetos", Dissertação de Mestrado, Universidade Federal do Vale do São Francisco, Juazeiro, BA, 2016.
- [13] F. A. Ferreira, T. S. Soares e A. C. G. C. Silva, "Análise da capacidade produtiva através do estudo de tempos e da simulação computacional: estudo de caso em um packing house na cidade de Petrolina-PE," in *Anais do XXXVIII Encontro Nacional de Engenharia de Produção*, Maceió, Alagoas, Brasil, 16-19 out. 2018, pp. 1-15.
- [14] Abrafrutas, "Com frutas para o mundo, Petrolina é a melhor cidade para o agronegócio" [Abrafrutas.org](https://abrafrutas.org). Disponível em: <https://abrafrutas.org/2021/11/com-frutas-para-o-mundo-petrolina-e-a-melhor-cidade-para-o-agronegocio/>, Acessado em: Data de acesso: 20 de junho de 2023.
- [15] Ramos, R.P., et al. Título do artigo: Non-invasive setup for grape maturation classification using deep learning Título do periódico: *Journal of Scientific Food and Agriculture* Ano: 2020.
- [16] A. Kasimati, B. Espejo-García, N. Darra, and S. Fountas, "Predicting Grape Sugar Content under Quality Attributes Using Normalized Difference Vegetation Index Data and Automated Machine Learning," *Sensors*, vol. 22, no. 6, p. 3249, 2022.
- [17] F. Palacios, M.P. Diago, and J. Tardaguila, "A Non-Invasive Method Based on Computer Vision for Grapevine Cluster Compactness Assessment Using a Mobile Sensing Platform under Field Conditions," *Sensors (Basel)*, vol. 19, no. 17, Sep. 2, 2019.
- [18] Luo, W. Liu, Q. Lu, J. Wang, W. Wen, D. Yan, and Y. Tang, "Grape Berry Detection and Size Measurement Based on Edge Image Processing and Geometric Morphology," *Machines*, vol. 9, no. 10, p. 233, 2021.
- [19] S. M. Jaisakthi, P. Mirunalini, D. Thenmozhi and Vatsala, "Grape Leaf Disease Identification using Machine Learning Techniques," *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)*, Chennai, India, 2019, pp. 1-6.
- [20] E. Ivorra, A.J. Sánchez, J.G. Camarasa, M.P. Diago, J. Tardaguila, Assessment of grape cluster yield components based on 3D descriptors using stereo vision, *Food Control*, Volume 50, 2015.
- [21] E. Ivorra, A.J. Sánchez, J.G. Camarasa, M.P. Diago, J. Tardaguila, Corrigendum to "Assessment of grape cluster yield components based on 3D descriptors using stereo vision" [*Food Control* 50 (April 2015) 273–282], *Food Control*, Volume 75, 2017.
- [22] G. Victorino, C. Poblete-Echeverría, and C. M. Lopes, "A Multicultivar Approach for Grape Bunch Weight Estimation Using Image Analysis," *Horticulturae*, vol. 8, no. 3, p. 233, 2022.
- [23] C. Zhang, H. Ding, Q. Shi and Y. Wang, "Grape Cluster Real-Time Detection in Complex Natural Scenes Based on YOLOv5s Deep Learning Network," *Agriculture*, vol. 12, p. 1242, 2022.
- [24] J. C. Mankins, "Technology Readiness Levels: A White Paper," Advanced Concepts Office, Office of Space Access and Technology, NASA, Apr. 6, 1995, edited Dec. 22, 2004.
- [25] [NIT-UNIVASF]. Nível de Maturidade Tecnológica (TRL/MRL), 2021.
- [26] Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação (PRPPGI), "Edital N° 09/2023 – Chamada Pública para o Programa de Jornada Empreendedora da INTECVASF", Universidade Federal do Vale do São Francisco, 2023.
- [27] SECTI - Secretaria de Ciência, Tecnologia e Inovação do Estado de Pernambuco. "Edital N° 09/2023 – Chamada Pública Para O Programa De Jornada Empreendedora Da Intecvasf", 2023
- [28] [NIT-UNIVASF]. Normas para o funcionamento da Incubadora de Empresas de Base Tecnológica da Universidade Federal do Vale do São Francisco, 2019.