KATNISS TRANSIT SEARCH: UM PIPELINE DE BAIXO CUSTO VIA INTELIGÊNCIA ARTIFICIAL PARA A VALIDAÇÃO DE CURVAS DE LUZ NO ENSINO DE FÍSICA COMPUTACIONAL

Maria Isabelly Lima de Sousa[1]

Bruno Victor Paiva da Silva [2]

Lino Martins de Holanda Júnior [3]

INTRODUÇÃO

Durante boa parte do século XX, a existência de planetas fora do Sistema Solar, denominados exoplanetas, permaneceu no campo das hipóteses. Embora sua existência fosse considerada provável, as limitações instrumentais dificultavam sua detecção. Entre as décadas de 1960 e 1980, diversas técnicas foram utilizadas nessa busca, incluindo astrometria, espectroscopia Doppler e fotometria. Algumas tentativas resultaram em falsas detecções, como a do suposto planeta da estrela de Barnard (Lubin et al., 2021). Em 1992, Aleksander Wolszczan e Dale Frail anunciaram a descoberta de planetas orbitando o pulsar PSR B1257+12, estabelecendo uma das primeiras confirmações de planetas fora do Sistema Solar (Wolszczan; Frail, 1992).

Essa descoberta demonstrou que planetas poderiam existir mesmo em ambientes extremos, como os sistemas de pulsares, formados após a evolução de estrelas massivas. Em 1995, a identificação de 51 Pegasi b, orbitando uma estrela semelhante ao Sol, representou outro marco importante para a área (Mayor; Queloz, 1995). Desde então, o

[1] Graduanda do Curso de Engenharia de Software da Universidade Federal Rural do Semi-Árido - UFERSA- RN,, maria.sousa05335@alunos.ufersa.edu.br;
[2] Pós Graduando em Engenharia Elétrica e de Computação pela Universidade Federal do Rio Grande do Norte - UFRN, bruno.silva.704@ufrn.edu.br
[3] Professor orientador: Doutor, Universidade Federal Rural do Semi-Árido - UFERSA - RN, lino.holanda@ufersa.edu.br

desenvolvimento de instrumentos e técnicas de análise ampliou significativamente o número de exoplanetas conhecidos.

Entre os métodos de detecção, destaca-se o trânsito planetário, que identifica a passagem de um planeta diante de sua estrela por meio de uma pequena redução no brilho observado. Esse método foi amplamente empregado em missões espaciais como Kepler e K2, da NASA, CoRoT, da Agência Espacial Europeia, e TESS, também da NASA (Borucki et al., 2010; Howell et al., 2014; Auvergne et al., 2009; Ricker et al., 2015). A grande quantidade de dados produzida por essas missões tornou necessário desenvolver ferramentas computacionais capazes de processar curvas de luz e identificar automaticamente possíveis sinais planetários.

Nesse contexto, Shallue e Vanderburg (2018) demonstraram o potencial das redes neurais convolucionais para classificar sinais de trânsito, distinguindo candidatos planetários de falsos positivos, como sistemas binários eclipsantes e artefatos instrumentais. A abordagem contribuiu para a identificação de Kepler-80g e Kepler-90i, sendo este último responsável por completar um sistema conhecido com oito planetas. Entretanto, o desenvolvimento desses modelos envolveu recursos computacionais expressivos, evidenciando a importância de investigar alternativas viáveis em ambientes com infraestrutura limitada.

Diante desse cenário, este trabalho tem como objetivo desenvolver e documentar um pipeline de processamento e classificação de curvas de luz da missão Kepler, combinando o algoritmo Transit Least Squares (TLS) com uma rede neural convolucional denominada KatnissNet. O pipeline realiza o pré-processamento dos dados, a extração de sinais periódicos e a geração de representações visuais e atributos estruturados, utilizados pelo modelo para distinguir registros de planetas confirmados e falsos positivos. A pesquisa utiliza dados previamente classificados do catálogo Kepler Object of Interest (KOI), não tendo como finalidade realizar uma busca cega e exaustiva por novos exoplanetas.

A pesquisa foi conduzida em um computador de recursos modestos, com processador Ryzen 5 e memória RAM limitada, utilizando dados públicos e bibliotecas gratuitas, como Lightkurve, NumPy e PyTorch. Além de avaliar o desempenho do modelo, o trabalho documenta as dificuldades de processamento e as estratégias de otimização adotadas. Dessa forma, busca demonstrar a viabilidade de desenvolver

atividades de análise astronômica e Física Computacional com baixo custo, contribuindo para ampliar o acesso a práticas científicas baseadas em dados.

METODOLOGIA

Embora o foco principal deste trabalho seja a aplicação de redes neurais na classificação de exoplanetas, foi necessário realizar etapas preliminares de processamento dos dados brutos. Para isso, utilizou-se o algoritmo TLS como ferramenta de pré-processamento de curvas de luz da missão Kepler. Diferentemente do que seria uma busca cega por novos planetas, o objetivo aqui foi exclusivamente preparar os dados, extrair os sinais periódicos, organizá-los e transformá-los em um formato adequado para alimentar futuros modelos de aprendizado de máquina.

Esta etapa permite demonstrar, na prática, como a astronomia contemporânea lida com grandes volumes de dados (big data), mesmo em computadores de baixo processamento. A escolha do TLS deve-se à sua maior sensibilidade a planetas de pequeno porte e à sua modelagem fisicamente realista dos trânsitos (Hippke; Heller, 2019), aspectos fundamentais para a geração de curvas de luz de qualidade.

FONTE E SELEÇÃO DE DADOS

Os dados foram obtidos a partir do catálogo Kepler Object of Interest (KOI), disponibilizado pelo NASA Exoplanet Archive (2026), no formato koi_template.csv. Foram considerados apenas objetos classificados como CONFIRMED e FALSE POSITIVE. As curvas de luz foram baixadas do MAST (Mikulski Archive for Space Telescopes) utilizando a biblioteca lightkurve.

Para garantir a confiabilidade do processamento, foram aplicados critérios mínimos de qualidade:

- Mínimo de 1.500 pontos válidos: curvas muito curtas dificultam a identificação de padrões periódicos.
- Cobertura temporal superior a 5%: fração do tempo total com medições válidas; valores muito baixos indicam muitos dados perdidos.

- Curva com variação significativa de fluxo: curvas sem variação (desvio padrão nulo) foram descartadas, pois geralmente indicam arquivos corrompidos ou estrelas constantes sem atividade.

PROCEDIMENTOS DE PROCESSAMENTO COM O TLS

O algoritmo TLS foi executado em ambiente Python 3.11, com bibliotecas lightkurve, numpy, pandas e astropy. As etapas de pré-processamento, com intuito de reduzir custo computacional, foram as seguintes:

- Definição da grade de busca: o intervalo de períodos foi definido dinamicamente. Quando o período orbital já era conhecido, a busca foi restrita a uma janela estreita, reduzindo o custo computacional.
- Remoção de tendências (flatten): variações de longo prazo foram removidas para que apenas os sinais transitórios permanecessem em destaque, utilizando uma janela de 501 pontos.
- Detecção e quantificação: a métrica principal foi a Signal Detection Efficiency (SDE). Adotaram-se dois limiares distintos: Busca cega (SDE > 5.0) e Busca guiada por período conhecido (SDE > 3.0).
- Geração das visualizações: para cada sinal detectado, foram geradas a Global view (curva dobrada em fase, normalizada) e a Local view (empilhamento dos trânsitos individuais centrados no evento).

AMBIENTE COMPUTACIONAL E ESTRATÉGIAS DE BAIXO PROCESSAMENTO

Os experimentos foram realizados em um computador de especificações modestas (processador Ryzen 5, 8 GB de RAM, 256 GB SSD). Para viabilizar o processamento dentro dessas limitações, adotaram-se as seguintes estratégias:

- Cache local: curvas de luz baixadas são armazenadas localmente, evitando downloads repetidos.

- Processamento sequencial em blocos: os dados são processados em pequenos lotes, distribuindo a carga computacional ao longo do tempo.
- Sistema de checkpoint: o progresso é salvo continuamente, permitindo interrupções e retomadas sem perda de dados.
- Uso de busca guiada: sempre que possível, o período conhecido é utilizado para restringir a grade de busca, reduzindo drasticamente o tempo de processamento.

No início dos experimentos, foi implementado um esquema de paralelismo para acelerar o processamento. Contudo, observou-se que, em um computador doméstico, essa abordagem resultava em maior tempo de execução e sobrecarga de memória, tornando o sistema instável. Por esse motivo, optou-se por manter o processamento sequencial em blocos, que se mostrou mais eficiente e compatível com o hardware disponível. Esse resultado também pode ser discutido em sala de aula, mostrando que soluções sofisticadas nem sempre são as mais adequadas em ambientes de baixo custo, e permitindo explorar com os alunos a relação entre algoritmos, desempenho e limitações físicas do equipamento.

PARÂMETROS UTILIZADOS NO TLS

Os principais parâmetros configurados foram:

- SDE (Signal Detection Efficiency): métrica central que quantifica a significância do sinal em relação ao ruído.
- Fator de superamostragem = 3: controla a resolução da grade de períodos; valores mais altos aumentam a sensibilidade, mas também o custo computacional.
- Profundidade mínima de trânsito = 10^{-5}: permite detectar sinais rasos, típicos de planetas pequenos.
- Número mínimo de trânsitos = 3: evita detecções espúrias por flutuações aleatórias.
- Passo da grade de duração = 1,1: mantém precisão relativa constante para trânsitos curtos e longos.

Todos os parâmetros foram ajustados empiricamente ao longo de meses de testes, buscando um equilíbrio entre sensibilidade, especificidade e viabilidade computacional em hardware modesto.

Estas estratégias demonstram que é possível trabalhar com big data astronômico mesmo sem acesso a supercomputadores, desde que haja planejamento e escolhas algorítmicas adequadas.

LIMITAÇÕES E ESCOPO

É importante destacar que este pré-processamento não constitui uma busca exaustiva por novos exoplanetas. Mais do que isso, queremos deixar explícito que esta pesquisa enfrentou, e documentou diversos obstáculos que normalmente não aparecem em artigos científicos tradicionais, mas que são parte essencial da prática científica.

As principais limitações observadas foram:

- Restrições de memória RAM: curvas de luz muito extensas causaram lentidão e, em alguns casos, falhas no processamento, exigindo a redução da grade de busca.
- Processamento interrompido: por utilizar um computador de uso compartilhado (estudos, navegação, outras tarefas), o fluxo de processamento foi interrompido diversas vezes. O sistema de checkpoint implementado mitigou o problema, mas não o eliminou completamente.
- Ajuste empírico de parâmetros: diversos parâmetros foram ajustados por tentativa e erro ao longo de meses. Uma das tentativas foi o uso de agregação temporal (binning) como possível solução para reduzir o custo computacional, conforme sugerido em alguns contextos. No entanto, na prática, essa abordagem mostrou-se irrelevante para os resultados obtidos, além dos próprios criadores do TLS não recomendarem essa prática, já que o algoritmo foi feito para sinais sensíveis e essa agregação afetaria esses sinais.
- Curvas de baixa qualidade: lacunas significativas nos dados geraram global views e local views com formas irregulares, que dificultariam a classificação por redes neurais.

- Impossibilidade de validação cruzada extensiva: devido às limitações de tempo e processamento, não foi possível realizar uma validação cruzada completa dos parâmetros otimizados.

Apesar dessas limitações, o fluxo estabelecido foi suficiente para gerar um conjunto de curvas de luz processadas (global views e local views) que serviram como insumo para a etapa seguinte: a classificação automática via redes neurais convolucionais. A mensagem central não é a perfeição do pipeline, mas sim sua viabilidade em condições reais e de baixo custo.

ARQUITETURA DA REDE NEURAL: KATNISSNET

A arquitetura da rede neural desenvolvida neste trabalho, denominada KatnissNet, fundamenta-se inicialmente em duas entradas principais: a global view (curva de luz completa dobrada em fase) e a local view (empilhamento dos trânsitos individuais). Ambas são processadas por ramos independentes de camadas convolucionais unidimensionais, que extraem características hierárquicas e morfológicas dos sinais.

Entretanto, devido a restrições computacionais durante a etapa de processamento para a criação da base de dados, constatou-se que nem todas as curvas de luz atingiram a qualidade ideal. A presença de lacunas temporais, ruído instrumental residual e artefatos oriundos do processo de flattening resultam em global views e local views degradadas. Tais irregularidades morfológicas poderiam comprometer severamente o desempenho do modelo caso a classificação dependesse exclusivamente dessas representações espaciais.

Para mitigar essas limitações, a arquitetura KatnissNet incorpora um terceiro fluxo de informação composto por atributos estruturados (features auxiliares). Esses atributos são provenientes de duas origens distintas:

- Atributos extraídos pelo TLS (Transit Least Squares) durante o pré-processamento: período orbital, Signal Detection Efficiency (SDE), duração do trânsito e odd-even mismatch (diferença de profundidade entre trânsitos pares e ímpares). Estas métricas são derivadas diretamente pelo algoritmo de busca sobre os dados recém-processados, garantindo alta consistência com as séries temporais inseridas no modelo.

- Atributos provenientes do catálogo KOI (Kepler Object of Interest): razão sinal-ruído do modelo ajustado (koi_model_snr), parâmetro de impacto (koi_impact), profundidade do trânsito em partes por milhão (koi_depth) e raio planetário estimado em raios terrestres (koi_prad). Estes valores foram obtidos do arquivo oficial disponível no NASA Exoplanet Archive.

Esses atributos atuam como um suporte informacional complementar, fornecendo à rede evidências adicionais para a discriminação do sinal. Do ponto de vista prático, enquanto os ramos convolucionais extraem as características morfológicas do trânsito a partir das curvas de luz, as features auxiliares fornecem os descritores físicos e estatísticos do candidato.

Como estratégia de regularização e robustez metodológica, aplicou-se uma técnica de Modality Dropout (exclusão de modalidade) durante a fase de treinamento. Em 20% dos batches, os vetores de features auxiliares são intencionalmente anulados (atribuídos a zero). Essa abordagem força a rede neural a extrair padrões discriminativos baseando-se unicamente nas curvas de luz, preparando-a para cenários de inferência onde os metadados possam estar ausentes ou corrompidos. Em contrapartida, quando os atributos estão presentes, a rede os utiliza para refinar suas predições, o que se mostra crucial para a classificação correta de amostras com curvas de luz de baixa qualidade.

PREPARAÇÃO DO CONJUNTO DE DADOS E DIVISÃO TREINO-TESTE

O conjunto de dados utilizado neste trabalho foi construído a partir da combinação de três fontes: (i) as curvas de luz processadas pelo TLS, (ii) as métricas extraídas pelo TLS (período, SDE, duração, odd-even mismatch), e (iii) os metadados do catálogo KOI (SNR, impacto, profundidade, raio planetário). O conjunto de dados final totalizou 6.833 exemplos com rótulos confiáveis. Desse total, 2.720 exemplos (39,8%) correspondem a planetas confirmados (CONFIRMED), enquanto 4.113 exemplos (60,2%) correspondem a falsos positivos (FALSE POSITIVE). Essa distribuição reflete a composição dos dados selecionados a partir do catálogo Kepler Object of Interest (KOI).

A divisão dos dados foi estratificada em dois subconjuntos mutuamente exclusivos: treinamento (80%), utilizado para ajustar os pesos da rede neural, e teste (20%), destinado à avaliação do desempenho do modelo. A divisão manteve a

proporção de planetas confirmados e falsos positivos em ambos os subconjuntos. Não foi utilizado um conjunto de validação separado. Durante o treinamento, a perda do conjunto de teste foi monitorada a cada época como critério de early stopping. Essa escolha representa uma limitação metodológica, pois o uso do conjunto de teste para orientar a interrupção do treinamento pode influenciar a avaliação final do modelo. Portanto, os resultados devem ser interpretados considerando essa limitação.

DETALHES DE IMPLEMENTAÇÃO E TREINAMENTO

A KatnissNet foi implementada em Python utilizando a biblioteca PyTorch, uma ferramenta gratuita e de código aberto. Todo o treinamento foi realizado no mesmo computador doméstico usado no pré-processamento (processador Ryzen 5, 8 GB de memória RAM, sem placa de vídeo dedicada).

Para o treinamento da rede, os pesos foram atualizados por meio do algoritmo de backpropagation associado ao otimizador Adam. Adotou-se uma taxa de aprendizado de 3×10^{-4} como estratégia para estabilizar o aprendizado da arquitetura em ambiente de hardware doméstico, mitigando oscilações divergentes na função de custo.

A função de perda mede o erro da rede. Utilizamos a Focal Loss, que dá menos importância para exemplos fáceis (trânsitos profundos ou falsos positivos óbvios) e foca nos casos difíceis (sinais fracos ou curvas ruidosas). Essa escolha é útil porque nosso conjunto tem muitos falsos positivos e poucos planetas confirmados.

O batch size define quantos exemplos a rede analisa antes de cada ajuste. Escolhemos 32, ou seja, a rede analisa 32 curvas de luz, calcula o erro médio e então ajusta seus pesos.

Para evitar que a rede decore os exemplos em vez de aprender o padrão geral, utilizamos duas técnicas:

- Dropout de 30%: desliga neurônios aleatoriamente, forçando a rede a não depender de caminhos específicos;
- Weight decay de 1×10^{-4}: penaliza pesos muito altos, mantendo o modelo simples.
- O treinamento foi interrompido automaticamente quando a perda no conjunto de teste parou de melhorar por 12 épocas consecutivas (early stopping). Uma época

significa que a rede já viu todos os exemplos de treinamento uma vez. Foi definido 50 épocas como limite máximo, mas na prática o treinamento parou antes.

- Ao final, salva-se os três modelos com menor perda no conjunto de teste. O melhor deles foi selecionado para fazer predições nos candidatos ainda não classificados. Todo o código está disponível em um repositório público de código aberto: https://github.com/isabellylimals/katniss-data-processing.git.

Os principais hiperparâmetros utilizados estão resumidos abaixo:

- Otimizador: Adam: Algoritmo que ajusta os pesos da rede;
- Taxa de aprendizado: 3×10^{-4}: Tamanho do passo de cada ajuste;
- Função de perda: Focal Loss: Mede o erro, focando nos casos difíceis;
- Batch size: 32: Exemplos processados antes de cada ajuste;
- Dropout: 30% : Evita que a rede "decore" os dados;
- Weight decay: 1×10^{-4}: Penaliza pesos muito altos;
- Early stopping: 12 épocas: Interrompe o treinamento sem melhora;
- Épocas máximas: 50: Limite de segurança.

REFERENCIAL TEÓRICO

Nas últimas décadas, observações astronômicas e avanços instrumentais permitiram demonstrar que os exoplanetas são comuns em nossa galáxia, entretanto, na maioria dos casos, esses corpos celestes não podem ser detectados diretamente por meio de telescópios ópticos convencionais, devido às enormes distâncias envolvidas e ao fato de não emitirem luz própria, refletindo apenas uma pequena fração da luz proveniente de suas estrelas hospedeiras. Além disso, a luminosidade estelar é tipicamente milhões (ou até bilhões) de vezes mais intensa que a refletida pelos exoplanetas, produzindo um intenso efeito de ofuscamento que dificulta severamente sua observação direta. Como consequência, apenas uma pequena parcela dos exoplanetas conhecidos, pouco superior a 1%, foi detectada por imageamento direto. Assim, a grande maioria das descobertas exoplanetárias é realizada por meio de métodos indiretos, nos quais se analisam os

efeitos gravitacionais ou fotométricos que esses corpos exercem sobre suas estrelas hospedeiras, como variações no movimento orbital ou na intensidade luminosa observada.

Neste trabalho focaremos nossa atenção no método de detecção que proporcionou a maior quantidade de dados observacionais e, consequentemente, o maior número de exoplanetas confirmados: o método do trânsito planetário.

A FÍSICA DO TRÂNSITO PLANETÁRIO

Um trânsito planetário ocorre quando um planeta passa diretamente entre sua estrela hospedeira e um observador. Esse bloqueio diminui ligeiramente o brilho da estrela, gerando uma curva de luz em função do tempo, conforme ilustrado abaixo.



Figura 1: Representação de um trânsito planetário, onde um exoplaneta passa em frente à sua estrela hospedeira e bloqueia parte da luz emitida pela estrela. O resultado observado é a diminuição da intensidade da luz emitida pela estrela.

O início do trânsito ocorre quando o exoplaneta tem o primeiro contato com o disco estelar; nesse ponto verificamos um decréscimo do fluxo na curva de luz. Esse decréscimo é máximo quando o exoplaneta está totalmente na frente do disco estelar, o que continua até a fase final do trânsito, quando ele inicia a saída da frente do disco

estelar e o fluxo começa a aumentar, voltando ao patamar inicial quando finalizado o egresso (José; Artur; João, 2020).

Para determinar o tamanho do exoplaneta, os astrofísicos se baseiam em princípios termodinâmicos e radiométricos (como a aproximação de corpo negro e as leis de emissão de Planck e Stefan-Boltzmann) aplicados à luz estelar (Young; Freedman, 2009; Santos; Amorim, 2017). Ao analisar a temperatura superficial da estrela por meio de seu espectro, calcula-se a intensidade de sua radiação. Com a distância conhecida, obtém-se a potência estelar. Durante o evento do trânsito, observa-se a queda relativa dessa potência. Essa queda fracionária de luminosidade equivale à razão direta entre a área do exoplaneta obstrutor e a área visível do disco da estrela hospedeira (José; Artur; João, 2020; Santos; Amorim, 2017). Dessa maneira geométrica e radiométrica, o raio do planeta pode ser inferido unicamente por meio da luz captada na Terra.

Além do tamanho, outro parâmetro vital é o período orbital. Para planetas próximos às suas estrelas, os trânsitos se repetem em intervalos curtos, permitindo que o período seja extraído observacionalmente de forma direta. No entanto, para órbitas longas, os cientistas utilizam formulações oriundas da Terceira Lei de Kepler e considerações sobre a geometria do trânsito — relacionando a duração do evento e a velocidade orbital à massa da estrela. Assim, é possível determinar características orbitais precisas e modelar o sistema planetário sem depender da observação de trânsitos sucessivos ao longo de muitos anos.

O ALGORITMO TRANSIT LEAST SQUARE (TLS)

Como evolução natural das ferramentas de busca por trânsitos, o algoritmo TLS foi proposto em 2019 por Hippke e Heller com o objetivo de superar as limitações do clássico BLS, introduzido por Kovacs, Zucker e Mazeh (2002). O BLS assumia uma forma simplificada em "caixa" (boxcar), ignorando fenômenos astrofísicos essenciais, como as bordas gradativas de ingresso/egresso e o escurecimento de limbo estelar.

O TLS introduz um modelo fisicamente realista, operando sobre todos os pontos da curva de luz (unbinned) e preservando a morfologia fina, o que aumenta drasticamente a precisão na busca por análogos terrestres (Hippke; Heller, 2019). Ele

atua amostrando períodos orbitais de forma otimizada e restringindo intervalos de duração baseados nas densidades estelares e Leis de Kepler. A detecção baseia-se no Signal Residue (SR) e no Signal Detection Efficiency (SDE) para quantificar o ajuste e a probabilidade de um sinal. Em testes oficiais, o TLS chegou a recuperar 93% de planetas de dimensão terrestre inseridos sinteticamente, consagrando-se como uma ferramenta de alta sensibilidade (Hippke; Heller, 2019).

RESULTADOS E DISCUSSÃO

Nesta seção, sintetizamos os principais achados do pipeline KatnissNet no conjunto de teste, destacando a capacidade de generalização do modelo e a utilidade pedagógica da extração fotométrica.

A análise exploratória das características derivadas pelo TLS revela um padrão claro: falsos positivos tendem a apresentar *Signal Detection Efficiency* (SDE) mediano superior ao dos exoplanetas confirmados em regimes de períodos orbitais ultra-curtos (< 10 dias), comportamento associado às intensas assinaturas de binárias eclipsantes. Em contrapartida, para períodos orbitais superiores a 50 dias, o SDE dos exoplanetas supera o dos falsos positivos, reforçando a importância do período orbital como parâmetro discriminativo.

A Tabela 1 sumariza as métricas operacionais obtidas pela KatnissNet no conjunto de teste, composto por 1.367 candidatos. Ressalta-se que, embora esses dados não tenham sido utilizados diretamente para atualizar os pesos da rede, sua perda foi monitorada durante o treinamento para definir o critério de early stopping.

Tabela 1 – Resumo do desempenho preditivo e métricas medianas do pipeline.

Categoria / Métrica	Planetas Confirmados	Falsos Positivos
SDE Mediano (Global)	34,5	41,5
Duração do Trânsito (Mediana)	0,169 dias (~4,0 h)	0,122 dias (~2,9 h)

Precisão da KatnissNet	0,75 (75%)	0,87 (87%)
Sensibilidade (Recall)	0,81 (81%)	0,82 (82%)

O modelo alcançou uma área sob a curva ROC (AUC-ROC) de 0,886, com acurácia global de 82%. A matriz de confusão correspondente é ilustrada abaixo:

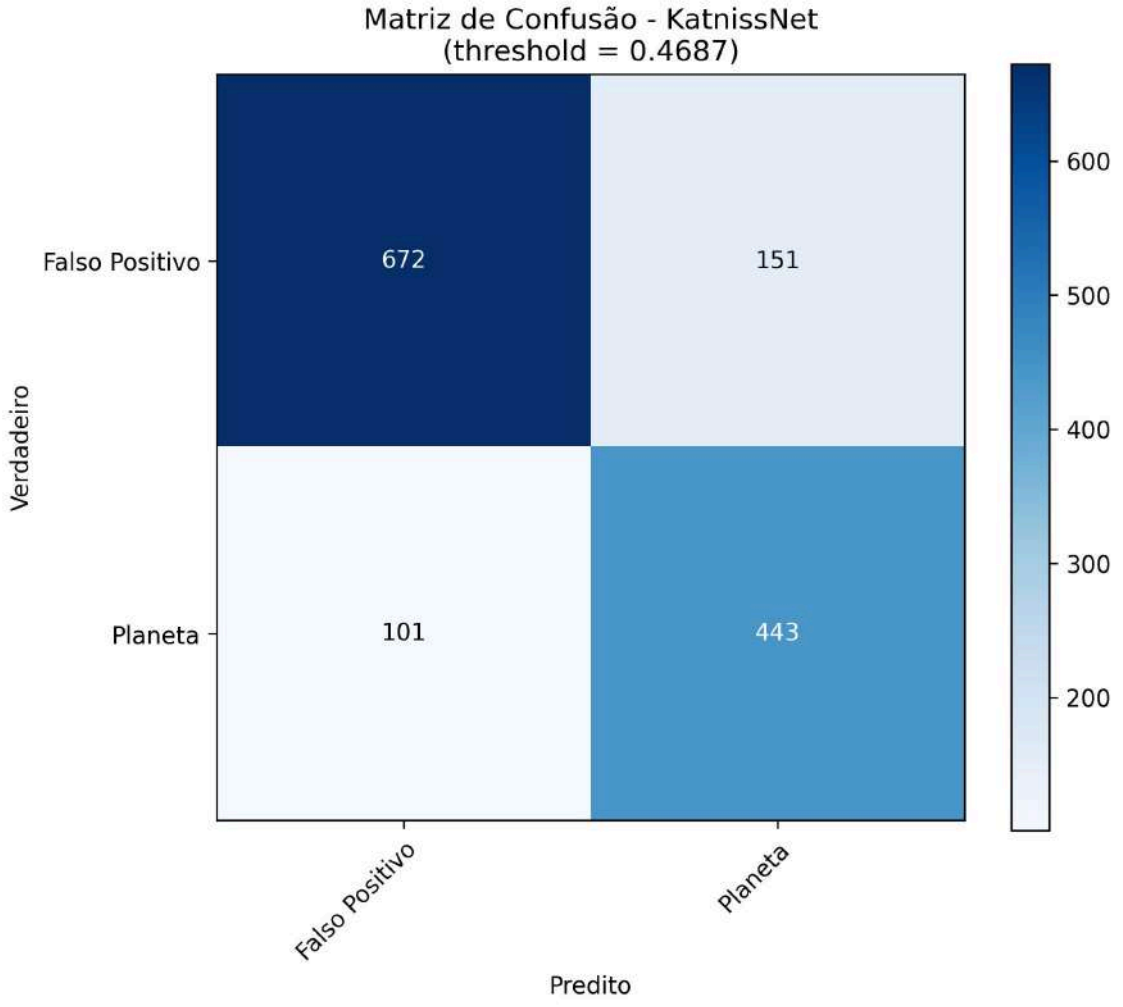


Figura 2: Matriz de confusão da KatnissNet obtida para o conjunto de teste. Os valores dispostos na diagonal principal indicam as classificações corretas.

A partir dos dados, computaram-se os parâmetros formais na Tabela 1. A arquitetura obteve acurácia global de 82%.

VISUALIZAÇÃO DE CURVAS DE LUZ PROCESSADAS

A capacidade da rede convolucional em discriminar padrões fotométricos é evidenciada no contraste das séries temporais dobradas em fase. O exoplaneta confirmado Kepler-2 b (HAT-P-7 b) (Pál et al., 2008) exibe o perfil simétrico suave em formato de "U" proveniente do escurecimento de limbo estelar. Em contraposição, o falso positivo K07010.01 exibe forte dispersão e assimetria, traços morfológicos característicos de sistemas binários de contato.

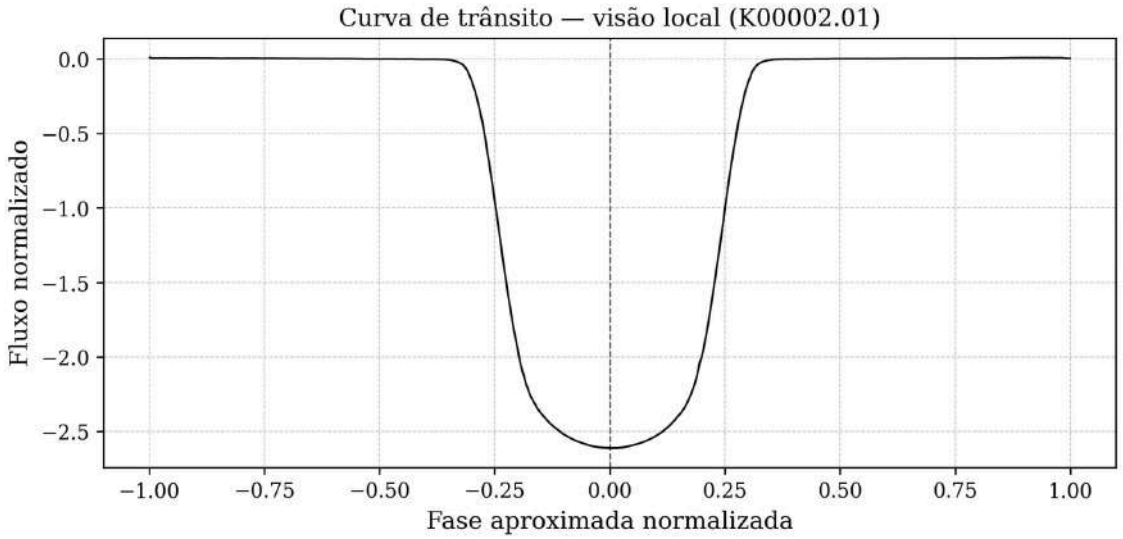


Figura 3: Curva de luz dobrada no período do exoplaneta confirmado Kepler-2 b (HAT-P-7 b), exibindo um perfil simétrico em "U"

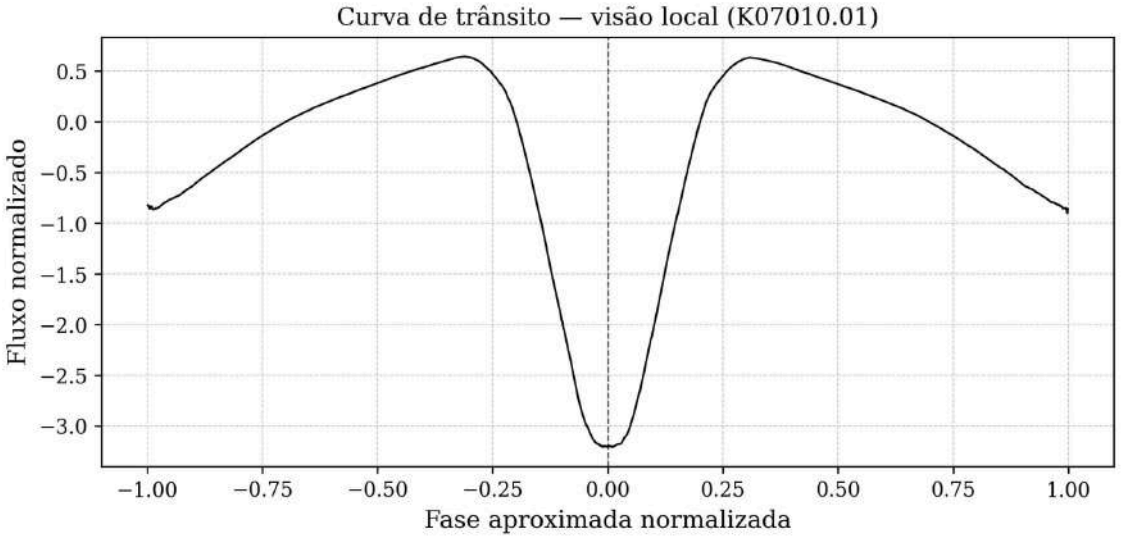


Figura 4: Curva de luz dobrada do falso positivo K07010.01, evidenciando dispersão e assimetria típicas de sistemas binários eclipsantes.

A extração automatizada desses recursos confirma que redes neurais enxutas capturam limites sutis de separabilidade em matrizes temporais ruidosas, dispensando a necessidade de filtros computacionais excessivamente pesados.

CONSIDERAÇÕES FINAIS

Os dados empíricos levantados sustentam a viabilidade de implementação de pipelines funcionais voltados à análise exoplanetária utilizando infraestruturas computacionais modestas. Os índices de acurácia (82%) e AUC-ROC (0,886) demonstram um desempenho satisfatório, mesmo sob severa restrição física de hardware (uso de estação doméstica sem aceleração por GPUs), mantendo-se a capacidade de lidar com ruído (curvas não perfeitamente limpas) e operando em regime desbalanceado (fiel à natureza dos dados da missão Kepler).

Embora arquiteturas industriais como a AstroNet (Shallue; Vanderburg, 2018) atinjam limiares de excelência como 0,988 de AUC-ROC valendo-se de processamentos massivos em 100 CPUs paralelas, o escopo do presente trabalho reside na demonstração de que contribuições científicas robustas podem ser articuladas com recursos mínimos. Esta característica viabiliza a inserção prática e a reprodução da pesquisa astrofísica

moderna e do ensino de física computacional em instituições com orçamentos restritos. Adicionalmente, a exposição sistematizada dos contratempos metodológicos atuou como um recurso didático valioso para os discentes e pesquisadores envolvidos, cumprindo o duplo papel analítico e formativo do pipeline.

Como limitação metodológica, destaca-se a ausência de um conjunto de validação independente, uma vez que a perda do conjunto de teste foi utilizada como critério de early stopping. Essa escolha pode influenciar a estimativa do desempenho final, indicando a importância de futuras avaliações com uma divisão independente em treinamento, validação e teste.

Palavras-chave: Exoplanetas; Física Computacional; Inteligência Artificial; Curvas de Luz; Ensino de Física.

REFERÊNCIAS

AUVERGNE, M. et al. The CoRoT satellite in flight: description and performance. **ArXiv:0901.2206**, 2009.

BORUCKI, William J. et al. Kepler Planet-Detection Mission: Introduction and First Results. **Science**, v. 327, 2010.

HIPPKE, Michael; HELLER, René. Optimized transit detection algorithm to search for periodic transits of small planets. **Astronomy & Astrophysics**, v. 623, n. A39, p. 1-13, 2019. DOI: 10.1051/0004-6361/201834672.

HOWELL, S. B. et al. The K2 Mission: Characterization and Early Results. **ArXiv:1402.5163**, 2014.

JOSÉ, C. S.; ARTUR, J. R.; JOÃO, C. P. Detecção do trânsito planetário de um exoplaneta com um telescópio de pequena abertura. **Revista Brasileira de Ensino de Física**, v. 42, 2020. DOI: https://doi.org/10.1590/1806-9126-RBEF-2020-0131.

KOVACS, G.; ZUCKER, S.; MAZEH, T. A box-fitting algorithm in the search for periodic transits. **Astronomy & Astrophysics**, v. 391, n. 1, p. 369-377, 2002. DOI: 10.1051/0004-6361:20020802.

LUBIN, Jack et al. Stellar Activity Manifesting at a One Year Alias Explains Barnard b as a False Positive. **arXiv:2105.07005**, 2021. DOI: https://arxiv.org/pdf/2105.07005.

MAYOR, Michel; QUELOZ, Didier. A Jupiter-mass companion to a solar-type star. **Nature**, v. 378, p. 355-359, 1995.

NASA EXOPLANET ARCHIVE. **NASA Exoplanet Archive**. Disponível em: https://exoplanetarchive.ipac.caltech.edu/. Acesso em: 30 mar. 2026.

PÁL, András et al. HAT-P-7b: An Extremely Hot Massive Planet Transiting a Bright Star in the Kepler Field. **The Astrophysical Journal**, v. 680, n. 2, p. 1450-1456, 2008. DOI: 10.1086/588010.

RICKER, G. R. et al. Transiting Exoplanet Survey Satellite (TESS). **Journal of Astronomical Telescopes, Instruments, and Systems**, v. 1, n. 014003, 2015.

SANTOS, W. C.; AMORIM, R. G. G. Descobertas de exoplanetas pelo método do trânsito. **Revista Brasileira de Ensino de Física**, v. 39, 2017. DOI: http://dx.doi.org/10.1590/1806-9126-RBEF-2016-0217.

SHALLUE, Christopher J.; VANDERBURG, Andrew. Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90. **The Astronomical Journal**, v. 155, n. 2, p. 94, 2018. DOI: 10.3847/1538-3881/aa9e09.

SIFFERT, Beatriz B. A procura por exoplanetas habitáveis. **Cadernos de Astronomia**, v. 3, n. 2, 2022.

WOLSZCZAN, A.; FRAIL, D. A. A planetary system around the millisecond pulsar PSR1257 + 12. **Nature**, v. 355, p. 145-147, 1992. DOI: https://doi.org/10.1038/355145a0.

YOUNG, H. D.; FREEDMAN, R. A. Física IV. Ótica e Física Moderna. **Pearson**, 2009.