Fundação Universidade Federal do ABC
Pró reitoria de pesquisa
Av. dos Estados, 5001, Santa Terezinha, Santo André/SP, CEP 09210-580
Bloco L, 3ºAndar, Fone (11) 3356-7617
iniciacao@ufabc.edu.br

Relatório Final de Iniciação Científica
referente ao Edital 2024

Nome do aluno: Thiago Fernandes Dias

Assinatura do aluno:

Nome do orientador: Paulo Henrique Pisani

Assinatura do orientador:

Título do projeto: Avaliação de abordagens de ajuste de hiperparâmetros em dinâmica de digitação.

Palavras-chave do projeto: biometria; aprendizado de máquina; ajuste de hiperparâmetros

Área do conhecimento do projeto: Ciências Exatas e da Terra; Ciência da Computação; Metodologias e Técnicas da Computação

Bolsista: Não

Santo André

2025

Resumo

O uso de biometria para autenticação de usuários, em oposição a métodos tradicionais baseados em senhas estáticas, têm sido atrativo em razão da maior segurança que um sistema biométrico pode proporcionar. Os usuários podem ser reconhecidos com base em características físicas ou comportamentais ao invés de precisarem se lembrar de uma senha ou a ter em mãos um cartão inteligente ou *token*. Dentre as diversas modalidades biométricas existentes, há a *dinâmica de digitação*, que reconhece as pessoas com base em seu ritmo de digitação. Diversos algoritmos de classificação podem ser usados nesse contexto. Esses algoritmos possuem hiperparâmetros, que precisam ser ajustados adequadamente. O ajuste de hiperparâmetros pode ser realizado para cada usuário individualmente ou pode ser global, isso é, os hiperparâmetros assumirão o mesmo valor para todos os usuários. O objetivo deste projeto é explorar diferentes abordagens de ajuste de hiperparâmetros em algoritmos para classificação para dinâmica de digitação.

Sumário

Lista de ilustrações

1 Introdução

Por volta dos anos de 1970, conforme explica Jain, Flynn e Ross (2007), engenheiros da IBM sugeriram que usuários de sistemas computacionais poderiam ser autenticados com base em suas características físicas ou comportamentais, únicas para cada pessoa. Com isso, seria possível limitar o acesso a informação somente a pessoas autorizadas de forma segura, sem a necessidade de a pessoa possuir algum tipo de cartão ou *token* de autenticação ou ter de memorizar uma senha, que poderia ser roubada por diversos meios.

Em (JAIN; FLYNN; ROSS, 2007) são descritos detalhadamente uma gama de modalidades biométricas físicas, como impressões digitais, padrão de veias nas mãos e íris, e comportamentais, como a caligrafia ou assinatura feita à mão. Cada uma possui a suas vantagens e desvantagens, além de disversas aplicações nos sistemas de autenticação atuais. Neste projeto, será dado o enfoque na vertente comportamental. Em particular, na *dinâmica de digitação* (ROY et al., 2022), que é uma modalidade biométrica que envolve analisar a forma com que um indivíduo digita em um teclado (MONROSE; RUBIN, 2000). Os usuários são, então, reconhecidos com base no seu ritmo de digitação.

Conforme definido em (JAIN; ROSS; PRABHAKAR, 2004), sistemas biométricos são sistemas de reconhecimento de padrões que extraem características de dados biométricos e então comparam as características extraídas com uma referência biométrica em um banco de dados. Essa comparação das características extraídas com a referência biométrica no banco de dados frequentemente resulta em uma pontuação (*score*) (JAIN; NANDAKUMAR; ROSS, 2016). Assumindo que seja uma pontuação indicando a similaridade, a classificação pode ser realizada aplicando um limiar de corte (*threshold*). Se a pontuação for maior que o limiar, o dado biométrico é classificado como genuíno e, caso contrário, como sendo de um impostor.

Nesse contexto, o limiar de corte pode ser entendido com um hiperparâmetro. O trabalho de Bischl et al. (2023) menciona o limiar de corte como um hiperparâmetro em algoritmos que retornam uma pontuação ou uma probabilidade. Outros algoritmos usados para reconhecimento de usuários pela dinâmica da digitação podem possuir outros hiperparâmetros para serem ajustados. Os valores dos hiperparâmetros tem um grande

impacto no desempenho preditivo de um sistema biométrico.

O objetivo deste projeto é comparar diferentes abordagens para ajuste de hiperparâmetros em algoritmos de classificação para dinâmica de digitação. Sobre esse aspecto, há algumas questões que podem ser investigadas. A primeira é sobre realizar o ajuste de forma global ou individualizada para cada usuário no sistema biométrico. Em biometria, o limiar de corte e a configuração de hiperparâmetros pode ser global, isso é, comum para todos os usuários, ou uma configuração específica para cada usuário (GIOT et al., 2011; MHENNI et al., 2019).

As demais seções do projeto estão organizadas da seguinte forma: na Seção 2, são introduzidos conceitos sobre dinâmica da digitação e ajuste de hiperparâmetros; na Seção 3 são descritos os conjuntos de dados utilizados e suas particularidades e os experimentos realizados; na Seção 4 são expostos os resultados e é feita uma discussão sobre eles e; na Seção 5 é feita a conclusão do projeto

2 Fundamentação Teórica

Esta seção apresenta alguns conceitos importantes para este projeto de pesquisa envolvendo dinâmica da digitação e ajuste de hiperparâmetros.

2.1 Dinâmica da digitação

A dinâmica da digitação é um modalidade biométrica comportamental que diferencia os indivíduos com base em atributos característicos da digitação de textos, como o tempo em que o indivíduo permanece pressionando cada tecla, o intervalo de tempo entre cada ativação de tecla e padrões nos erros de digitação (KARNAN; AKILA; KRISHNARAJ, 2011).

Segundo Peacock, Ke e Wilkerson (2004), a principal vantagem da dinâmica da digitação em relação às demais modalidades biométricas comportamentais é a sua transparência. Por exemplo, se usada em conjunto com um formulário de autenticação comum, composto por um identificador de usuário (ou nome de usuário, *username*) e uma senha, as métricas

de digitação podem ser obtidas das informações que o indivíduo necessariamente deverá inserir no sistema que ele deseja acessar. Além disso, em serviços baseados na Web, muitas vezes não é viável exigir formas de autenticação por biometria, pois os usuários podem não ter acesso aos dispositivos necessários, como câmeras e sensores de impressões digitais, ou equipamentos mais sofisticados.

Para fins de reconhecimento biométrico, os dados podem ser obtidos por meio dos padrões de digitação tanto de *textos fixos* definidos previamente, que os usuários serão requisitados a digitar para fins de identificação ou verificação, ou de *textos livres*, sem um tamanho fixo ou qualquer outra restrição. Muitos dos estudos desenvolvidos sobre autenticação pela dinâmica de digitação consideram um mecanismo baseado em texto fixo, geralmente o nome de usuário e a senha coletados previamente. Entretanto, diversos pesquisadores também aplicam algoritmos de aprendizado de máquina para desenvolver modelos capazes de autenticar os usuários de forma contínua por meio do texto digitado em um sistema durante o seu uso (LU et al., 2020).

2.1.1 Coleta e modelagem de dados

Considerando uma situação em que o texto a ser digitado é fixo, para criar modelos de Aprendizado de Máquina capazes de realizar a verificação de usuários a partir da forma com que eles digitam o texto pré-determinado é necessário um processo de treinamento a partir dos dados coletados durante a digitação do texto. Além disso, esses dados devem ser modelados corretamente para que sejam utilizados nos algoritmos de Aprendizado de Máquina.

Enquanto o usuário digita o texto são registrados os intervalos de tempo de ativação entre as teclas, sendo eles (DIAS et al., 2023):

- *Hold time*, ou *Dwell time*: tempo em que o usuário permaneceu pressionando uma tecla após a ativação;
- *Down-Down*, ou *Flight time*: intervalo de tempo entre a ativação de duas teclas consecutivas;

- *Down-Up*: intervalo de tempo entre a ativação de uma tecla e a desativação da próxima tecla
- *Up-Down*: intervalo de tempo entre a desativação de uma tecla e ativação da próxima tecla;
- *Up-Up*: intervalo de tempo entre a desativação de duas teclas.

A cada vez que um participante digita o texto fixo, esses intervalos são registrados. Eles compõem um vetor, que serve como dado de entrada em algoritmos de Aprendizado de Máquina.

2.2 Hiperparâmetros

De acordo com Yang e Shami (2020), desenvolver um modelo de aprendizado de máquina que seja efetivo na resolução de um determinado problema é uma tarefa complexa e demorada. Ela envolve a escolha do algoritmo apropriado e a obtenção de um modelo arquitetural ótimo por meio do ajuste de hiperparâmetros. O autor explica que há dois tipos de parâmetros em modelos de aprendizado de máquina: os *parâmetros* do próprio modelo, que serão inicializados e repetidamente atualizados durante o processo de treinamento, e os chamados *hiperparâmetros*, que devem ser escolhidos antes de o modelo ser treinado. Os autores mencionam os pesos dos neurônios em redes neurais como um exemplo de parâmetro de modelo e o parâmetro de penalidade C em uma *Support Vector Machine* (SVN), a taxa de aprendizado em redes neurais e o algoritmo utilizado para minimizar a função objetivo como hiperparâmetros.

O ajuste de hiperparâmetros é o processo de testar valores diferentes para os hiperparâmetros a fim de se obter o melhor ajuste para um modelo construído a partir de determinada base de dados. Na perspectiva de Hutter, Kotthoff e Vanschoren (2019), o ajuste de hiperparâmetros é uma parte fundamental da construção de modelos de aprendizado de máquina efetivos, especialmente em redes neurais artificiais e modelos baseados em árvores de decisão, que possuem diversos hiperparâmetros. DeCastro-García et al. (2019) explica que problemas de otimização de hiperparâmetros (*hyperparameter opti-*

mization, HPO) exigem um entendimento profundo da relação entre as combinações de hiperparâmetros e o modelo de aprendizado de máquina resultante do processo de treinamento. Ambos dependem do algoritmo utilizado e do tipo de cada hiperparâmetro, que pode ser contínuo, discreto ou categórico.

Segundo Yang e Shami (2020), após a escolha do algoritmo de aprendizado de máquina e dos métodos que serão utilizados para avaliar o seu desempenho, é necessário listar os hiperparâmetros que deverão ser ajustados e, então, definir os conjuntos de valores possíveis para cada um de acordo com o seu tipo. Dependendo do problema, os hiperparâmetros podem possuir restrições, isto é, não poderão assumir qualquer valor dentre todos os valores possíveis, e essas restrições impostas a um hiperparâmetro podem estar condicionadas aos valores escolhidos para outro. Além disso, para cada configuração diferente, o modelo deverá ser treinado e testado novamente, para que o seu desempenho seja medido.

Alguns motivos para aplicar técnicas de ajuste de hiperparâmetros em aprendizado de máquina são evitar a necessidade de realizar o ajuste manualmente, o aprimoramento do desempenho dos algoritmos, assim como a melhora da reprodutibilidade e justiça dos estudos realizados Hutter, Kotthoff e Vanschoren (2019), Bischl et al. (2023). A próxima seção discute algumas questões sobre o ajuste de hiperparâmetros no contexto de dinâmica da digitação, que será o foco deste projeto.

2.3 Ajuste de hiperparâmetros no contexto de dinâmica da digitação

Diversos trabalhos na área de dinâmica da digitação acabam não aplicando uma técnica de ajuste de hiperparâmetros em razão da métrica usada para reportar os resultados. Isso ocorre, por exemplo, ao reportar resultados em termos de EER (*Equal Error Rate*). Ao ajustar o limiar de corte de um sistema biométrico, as taxas de falsa aceitação (impostores aceitos erroneamente) e de falsa rejeição (usuários genuínos rejeitados de forma indevida) podem mudar. De maneira geral, ao aumentar uma taxa, a outra diminui dependendo do ajuste do limiar de corte. O valor EER representa o ajuste em que as duas taxas são iguais (ROY et al., 2022). Para isso, os rótulos de teste (genuíno/impostor) podem ser

usados para encontrar esse ajuste. Entretanto, em uma aplicação prática, o acesso aos rótulos dos dados pode não estar disponível.

Alguns trabalhos avaliaram o impacto dos hiperparâmetros. No trabalho elaborado por Purwar et al. (2019), a técnica *Grid Search* foi aplicada para encontrar a melhor combinação de hiperparâmetros para uma implementação do algoritmo *Support Vector Machine* (SVM) usado para reconhecimento de usuários pela dinâmica da digitação. Os estudos realizados por Kasprowski, Borowska e Harezlak (2022) avaliaram diferentes arquiteturas de redes neurais e a influência de hiperparâmetros como número de filtros convolucionais, tamanho do *kernel* de convolução, número de neurônios na camada recursiva e taxa de *drop out*.

Uma discussão sobre ajustar o limiar de corte de forma individual e de forma global foi realizada por Giot et al. (2011), assim como também avaliou a adaptação de modelos ao longo do tempo. De fato, em dinâmica da digitação, o ritmo de digitação pode mudar com o tempo. Outro trabalho que avaliou o ajuste de hiperparâmetros em dinâmica da digitação foi o de Mhenni et al. (2016). Nesse trabalho, foi considerado um cenário de sistemas biométricos adaptativos (RYU et al., 2023), em que a referência biométrica pode ser atualizada conforme os usuários realizam a autenticação. Os mesmos autores também discutiram essa adaptação do limitar de corte em (MHENNI et al., 2019).

3 Metodologia

Neste projeto foram comparadas técnicas para o ajuste de hiperparâmetros de algoritmos de classificação em dinâmica da digitação. Para isso, foram utilizados conjuntos de dados disponíveis publicamente, conforme descrito na Subseção 3.1. Esses dados foram divididos entre treino e teste, sendo que as amostras usadas para treinamento serão referentes a dados mais antigos em comparação com os dados usados para teste. Na Seção 3.2, são descritas métricas que serão usadas para avaliação de desempenho neste trabalho.

3.1 Conjuntos de dados

Grande parte dos trabalhos que realizaram experimentos com dados de dinâmica da digitação não disponibilizaram os dados coletados (ROY et al., 2022). Esse fato dificulta a reprodutibilidade de estudos na área. Neste projeto foram utilizados dados publicamente disponíveis. Os conjuntos de dados usados são descritos a seguir:

3.1.1 CMU

Este conjunto de dados [1] possui dados de 51 indivíduos que digitaram a senha ".tie5Roanl" em oito sessões de captura, com 50 amostras em cada sessão. No total, cada indivíduo digitou a senha 400 vezes. É importante destacar que, assim como em outros conjuntos de dados públicos e privados, houve um intervalo de tempo entre as sessões de coleta, para que as variações na digitação de cada usuário fossem consideradas no modelo. Cada pessoa participou de somente uma sessão por dia.

Os autores explicam que dentre os motivos da senha ser fixa para todos os usuários estão o viés que poderia surgir nos experimento se cada usuário pudesse escolher a própria senha e a necessidade de coletar amostras de impostores específicas para cada usuário, o que tornaria a coleta de dados ainda mais difícil.

3.1.2 KeyRecs

O conjunto de dados KeyRecs [2] envolveu a captura de dinâmica da digitação de texto fixo e de texto livre. Como o foco deste projeto foi em texto fixo, apenas essa parte do conjunto de dados foi utilizada. Para texto fixo, de acordo com a descrição do conjunto de dados, 99 indivíduos digitaram uma mesma senha em duas sessões, com 100 amostras em cada sessão, totalizando 200 amostras por indivíduo. Ao realizar o download da versão disponível, entretanto, observou-se que alguns usuários tem menos do que 200 amostras.

[1] <https://www.cs.cmu.edu/~keystroke/>
[2] <https://zenodo.org/records/7886743>

3.2 Métricas

Esta seção descreve algumas métricas usadas na literatura que serão usadas para avaliação dos resultados nos experimentos realizados neste projeto de pesquisa. Essas métricas são: FMR, FNMR e acurácia balanceada (Precise Biometrics, 2014; FERLINI et al., 2021). Uma breve descrição dessas métricas é apresentadas a seguir:

- FMR (*False Match Rate*, Taxa de falsa correspondência): percentual de tentativas de impostores que foram aceitas como genuínas, definida como

$$FMR = \frac{numero\ de\ tentativas\ de\ impostores\ aceitas}{total\ de\ tentativas\ de\ impostores}. \quad (1)$$

  Uma taxa relacionada é a FAR (*False Acceptance Rate*), que tem significado similar, mas considera também taxa em que o sistema biométrico falha ao obter uma amostra biométrica. Essa taxa é conhecida como FTA (*Failure to Acquire Rate*).

- FNMR (*False Non-match Rate*, Taxa de falsa não-correspondência): percentual de tentativas genuínas que foram rejeitadas como impostoras pelo sistema, definida como

$$FNMR = \frac{numero\ de\ tentativas\ genuinas\ rejeitadas}{total\ de\ tentativas\ de\ usuarios\ genuinos}. \quad (2)$$

  Uma métrica relacionada é a FRR (*False Rejection Rate*), que tem um significado similar, mas considera também a FTA.

- Acurácia balanceada: média do acerto para cada classe (genuíno e impostor). Essa métrica pode ser obtida a partir da o cálculo da (HTER - *Half Total Error*, Metade do erro total)

$$HTER = \frac{FNMR + FMR}{2}, \quad (3)$$

  definida como a média entre FNMR e FMR (ROY et al., 2022). A partir da HTER, então é obtida a acurácia balanceada, definida como

$$BAcc = 1 - HTER. \tag{4}$$

3.3 Divisão de dados de treino, validação e teste

A divisão de dados entre treino e teste foi feita com base no momento em que as amostras de digitação foram coletadas. Em ambos os conjuntos de dados foram utilizadas as primeiras 50 amostras de cada um dos usuários para treinamento, o que corresponde à primeira sessão de coleta do CMU e na metade dos registros da primeira sessão de coleta do Keyrecs. O restante dos dados de cada usuário foi utilizado para testes. No CMU, isso corresponde às 7 sessões que sucedem a primeira. No Keyrecs, isso correspondem à segunda metade das amostras da primeira sessão e todas as amostras da segunda sessão.

Em alguns algoritmos de Aprendizado de Máquina, somente os dados do usuário genuíno são utilizados para treinamento, e o modelo é então testado com dados de usuários impostores e de usuários genuínos. Estes modelos são chamados de *detectores de anomalias* (KILLOURHY; MAXION, 2009). Para outros algoritmos, ambos os dados de usuários impostores e genuínos devem ser utilizados para treinamento. Nesse sentido, é importante que a quantidade de amostras para cada classe (impostor e genuíno) sejam as mesmas, para que o modelo criado não seja enviesado. Dessa forma, foram selecionadas aleatoriamente 50 amostras de usuários impostores para treinamento, sendo que no CMU foi considerada apenas a primeira sessão de cada usuário impostor e no Keyrecs somente as primeiras 50 amostras para seleção.

Para o ajuste de hiperparâmetros foi feita uma validação cruzada sobre as amostras de treinamento, em que 4/5 do total de amostras foram utilizadas para a criação de um modelo com uma configuração de hiperparâmetros específica e 1/5 foi utilizada para testar o modelo. A média da $BAcc$ obtida em cada divisão foi considerada para selecionar a melhor configuração. Nos casos em que foi necessário utilizar dados de impostores para treinamento, os conjuntos de treinamento com amostras de usuários genuínos e impostores foram divididos separadamente e 4/5 de ambos os conjuntos foi utilizado para treinamento e 1/5 para testes.

3.4 Algorítmos e Hiperparâmetros

Foram utilizados três algoritmos de Aprendizado de Máquina neste trabalho, sendo eles o *Random Forest* (RF), a *Support Vector Machine* (SVM) e o algoritmo desenvolvido por Santos, Magalhães e Santos () (ST). Eles foram usados para criar modelos de verificação para cada usuário com base em suas amostras de digitação nos dois conjuntos de dados mencionados anteriormente. Com cada algoritmo foram avaliadas 3 abordagens:

- Sem ajuste de hiperparâmetros: os modelos foram criados com a configuração padrão de hiperparâmetros da biblioteca utilizada (Scikit-learn);

- Ajuste de hiperparâmetros por usuário: para cada usuário foram testadas todas as configurações possíveis de hiperparâmetros e foi selecionada a que gerou um modelo com maior BAcc;

- Ajuste de hiperparâmetros global: cada configuração de hiperparâmetros foi utilizada para a criação de um modelo para cada usuário e foi então calculada a BAcc média dos modelos. Foi selecionada a configuração que gerou modelos cuja a BAcc média foi a mais alta.

Em cada uma das abordagens a melhor configuração de hiperparâmetros encontrada foi utilizada para o treinamento dos modelos para cada usuário, com todo o conjunto de dados para treinamento.

4 Resultados e discussão dos resultados

Nesta sessão serão apresentados os resultados dos experimentos. Abaixo os gráficos de FMR/FNMR e BAcc para cada usuário, em ambos os conjuntos de dados utilizados.

Figura 1 – Magalhães (CMU) - FMR e FNMR

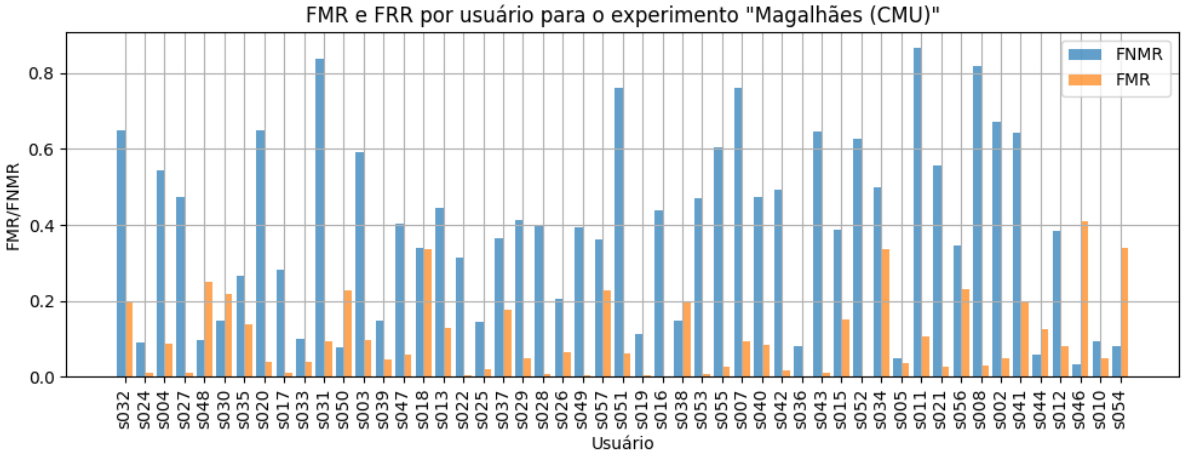


Figura 2 – Magalhães (Keyrecs) - FMR e FNMR

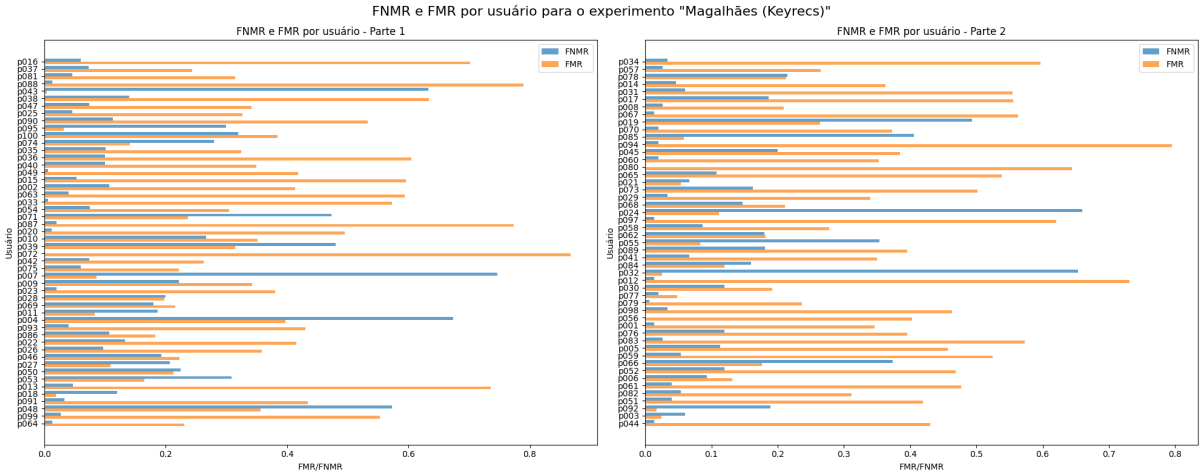


Figura 3 – Magalhães com HPO global (CMU) - FMR e FNMR

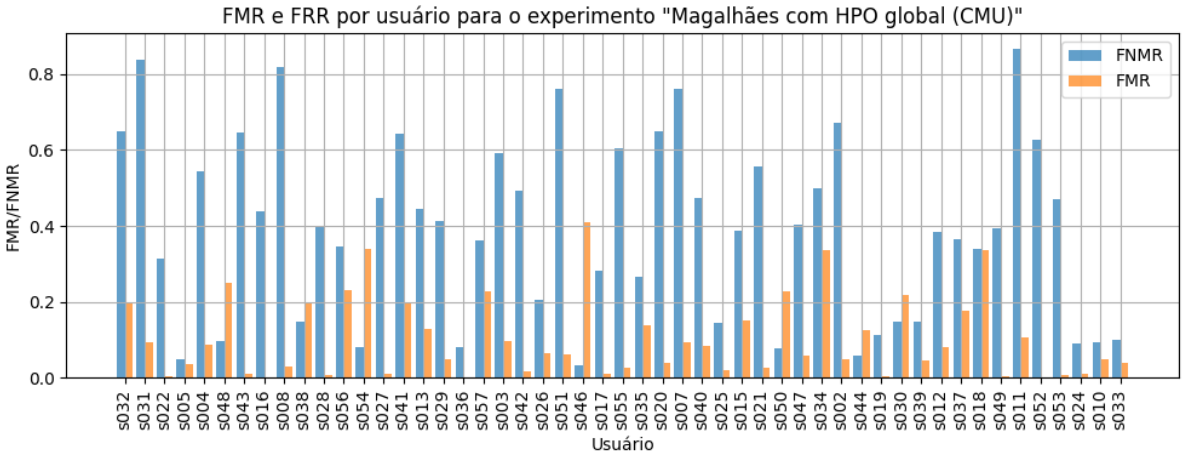


Figura 4 – Magalhães com HPO global (Keyrecs) - FMR e FNMR

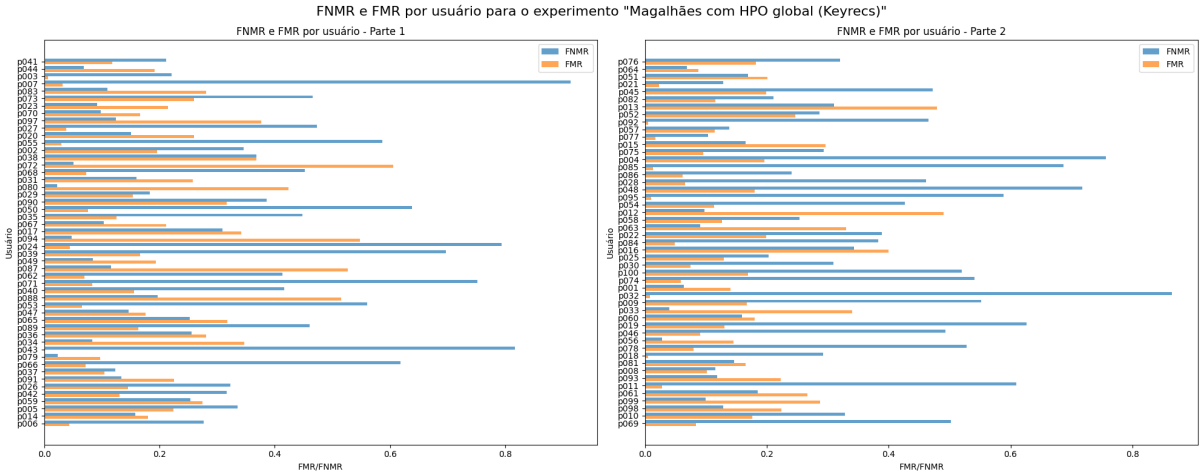


Figura 5 – Magalhães com HPO por usuário (CMU) - FMR e FNMR

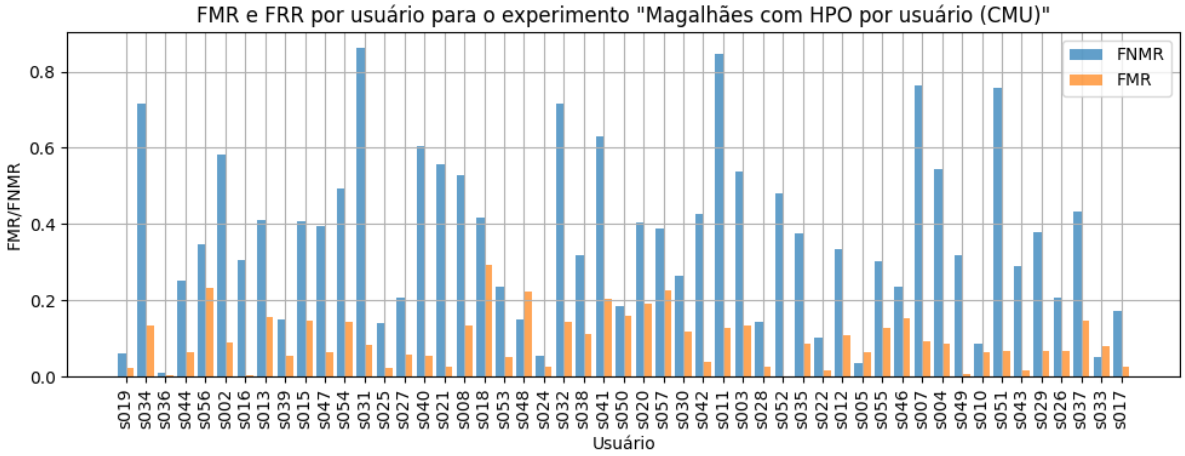


Figura 6 – Magalhães com HPO por usuário (Keyrecs) - FMR e FNMR

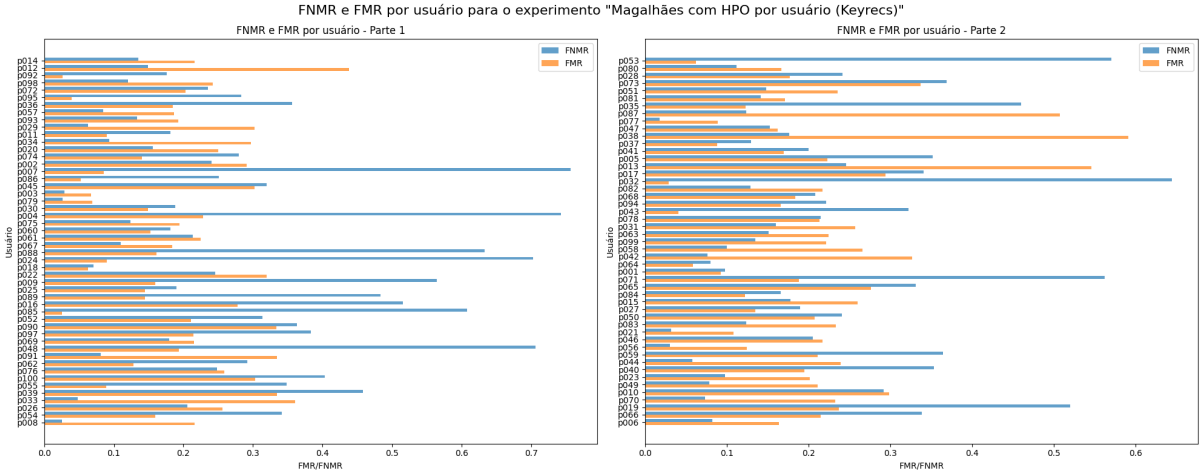


Figura 7 – Random Forest com HPO global (CMU) - FMR e FNMR

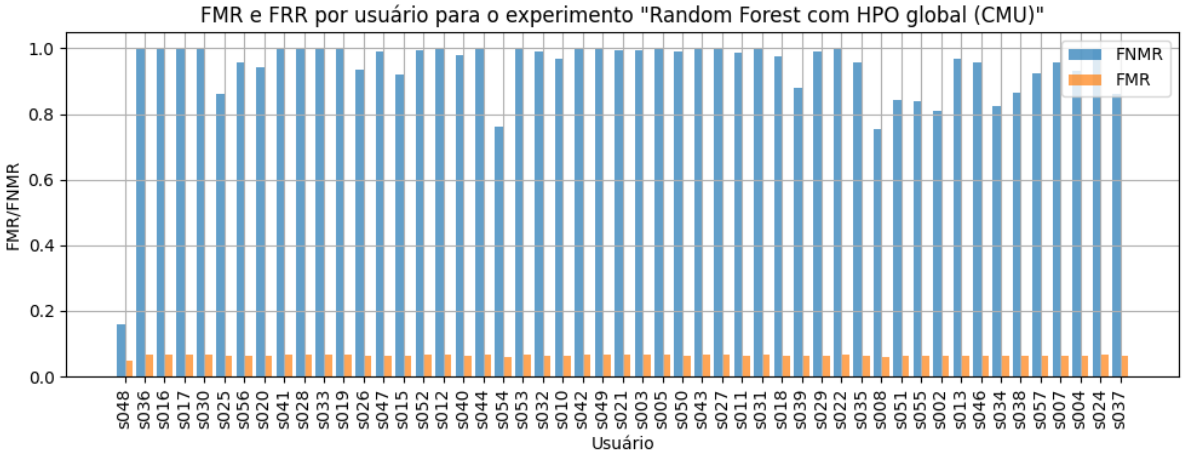


Figura 8 – Random Forest com HPO por usuário (CMU) - FMR e FNMR

Figura 9 – SVM (CMU) - FMR e FNMR

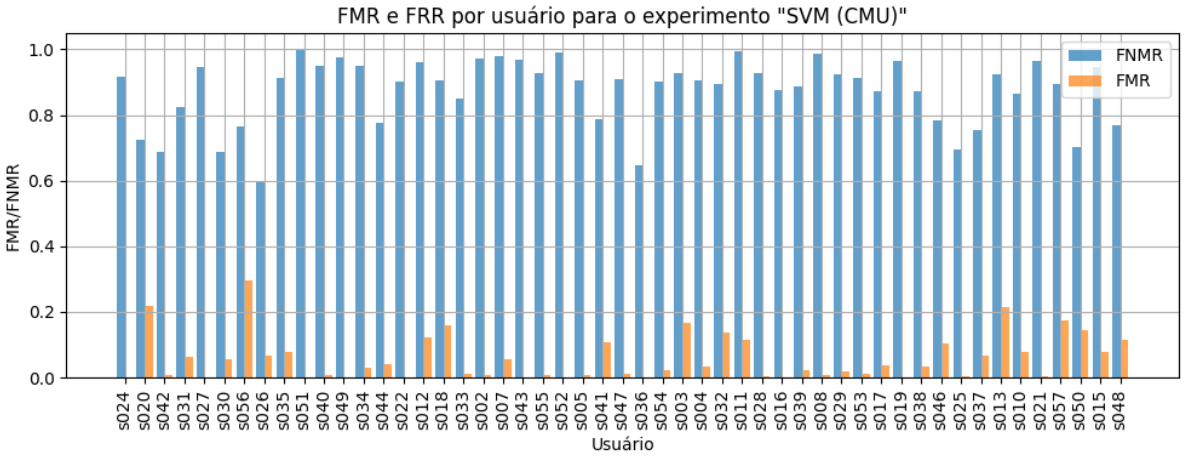


Figura 10 – SVM (Keyrecs) - FMR e FNMR

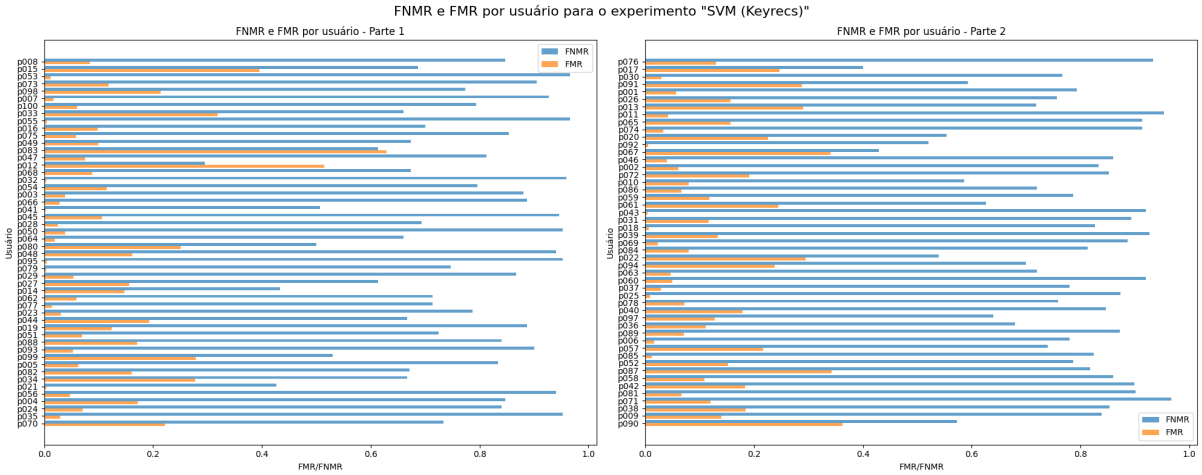


Figura 11 – SVM com HPO global (CMU) - FMR e FNMR

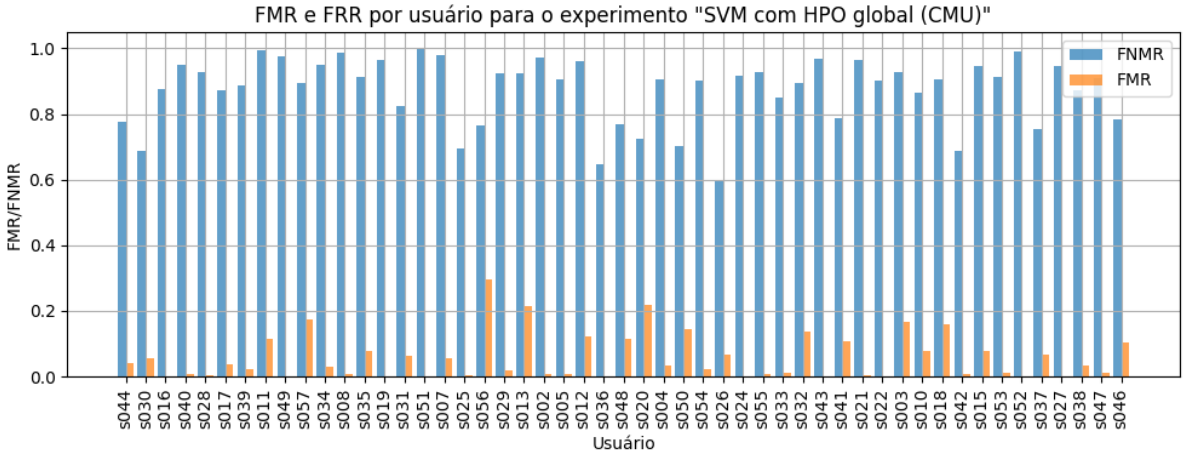


Figura 12 – SVM com HPO global (Keyrecs) - FMR e FNMR

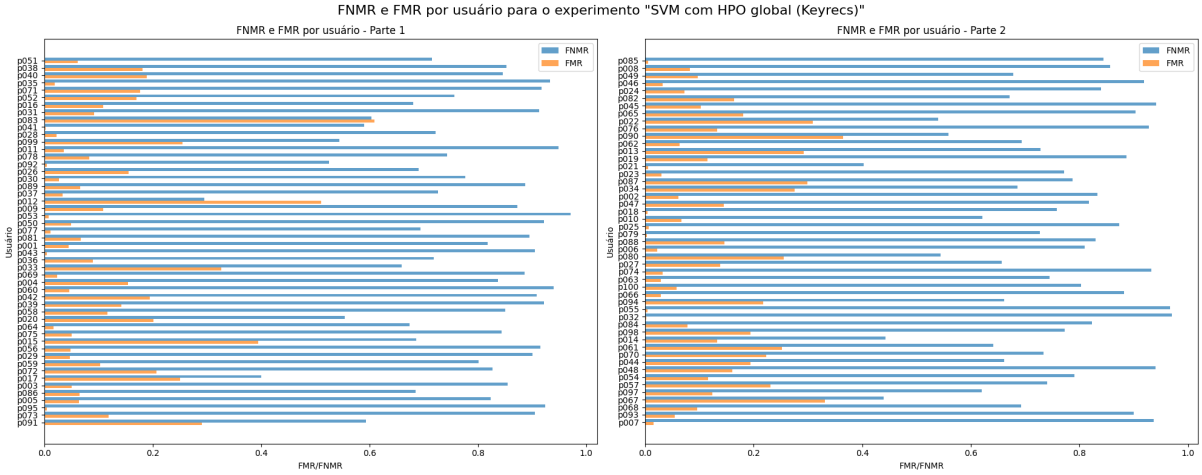


Figura 13 – SVM com HPO por usuário (CMU) - FMR e FNMR

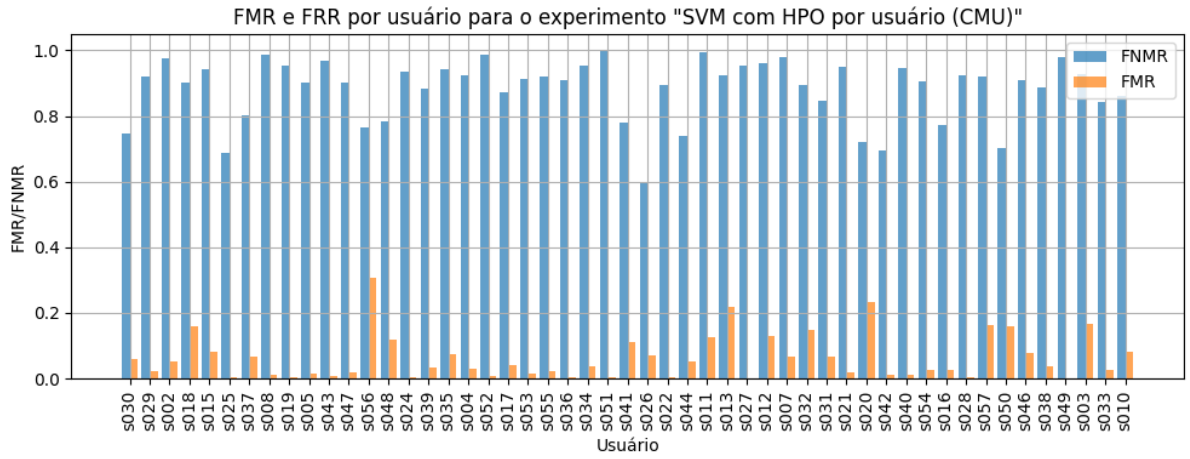


Figura 14 – SVM com HPO por usuário (Keyrecs) - FMR e FNMR

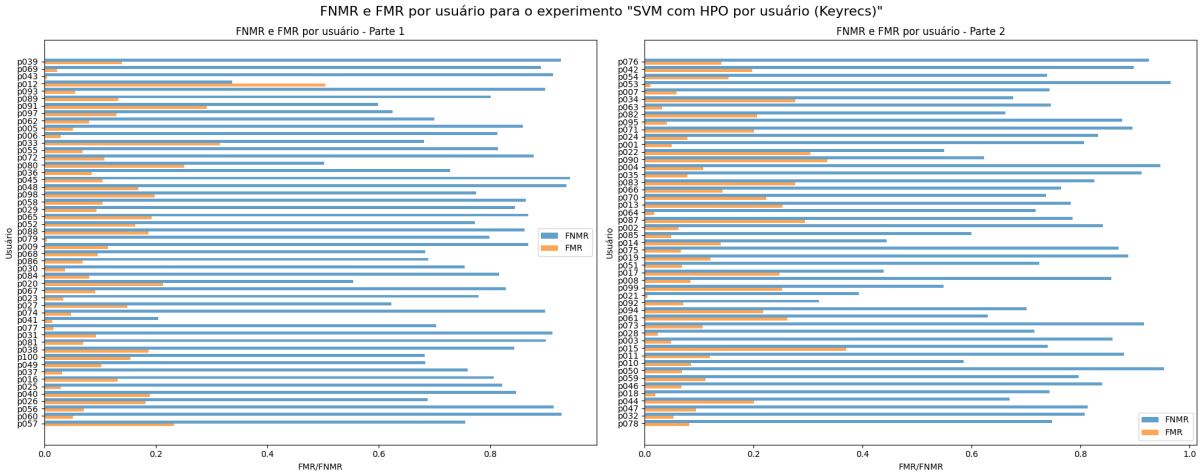


Figura 15 – Magalhães (CMU) - BAcc

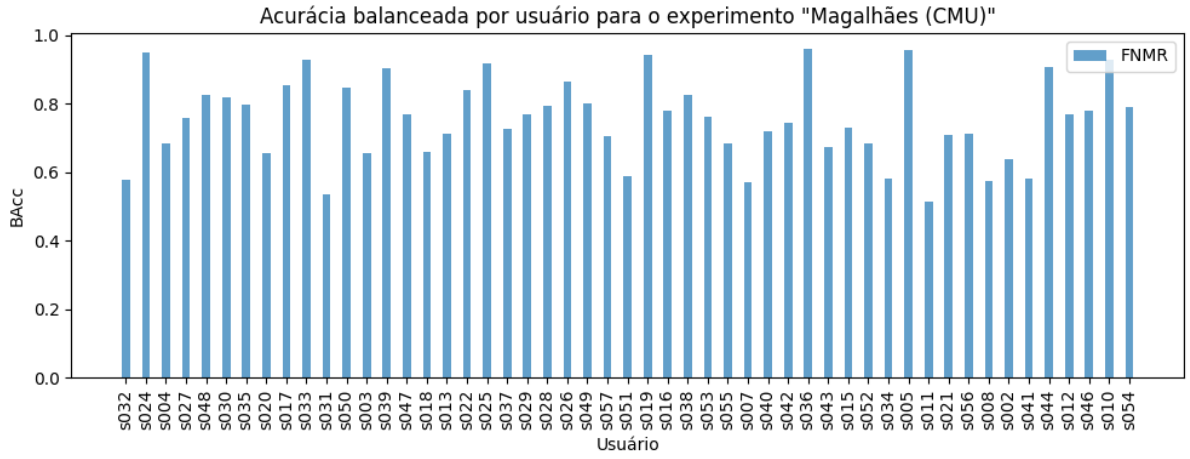


Figura 16 – Magalhães (Keyrecs) - BAcc

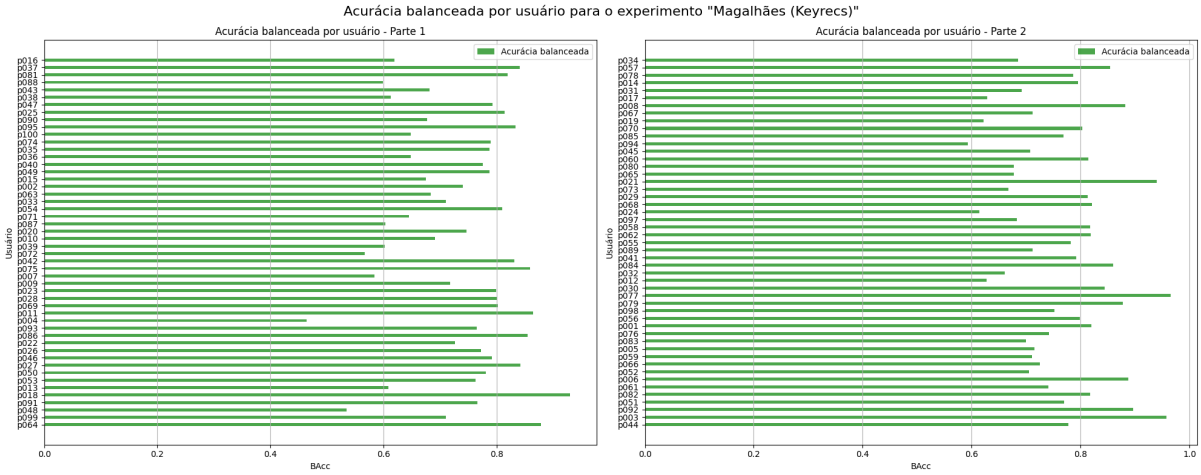


Figura 17 – Magalhães com HPO global (CMU) - BAcc

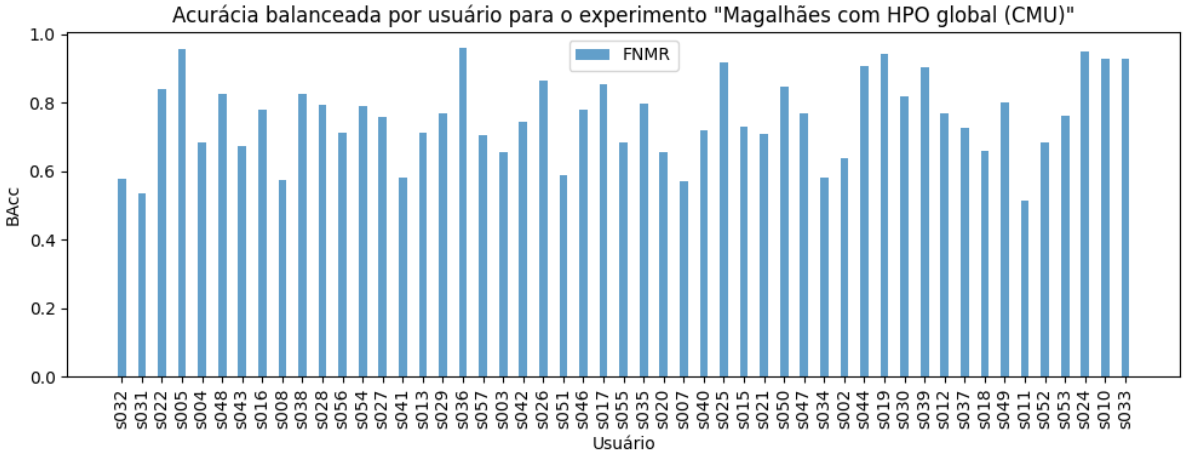


Figura 18 – Magalhães com HPO global (Keyrecs) - BAcc

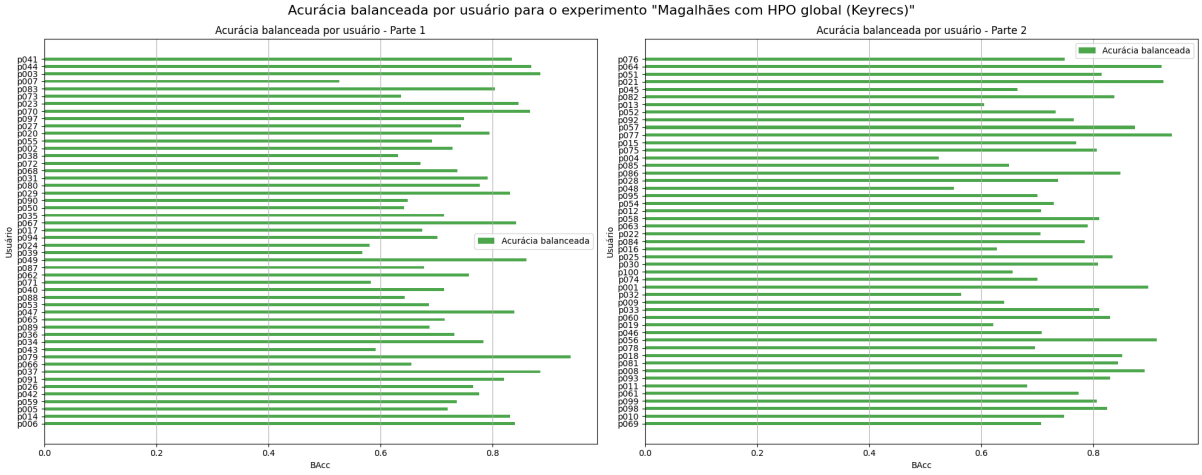


Figura 19 – Magalhães com HPO por usuário (CMU) - BAcc

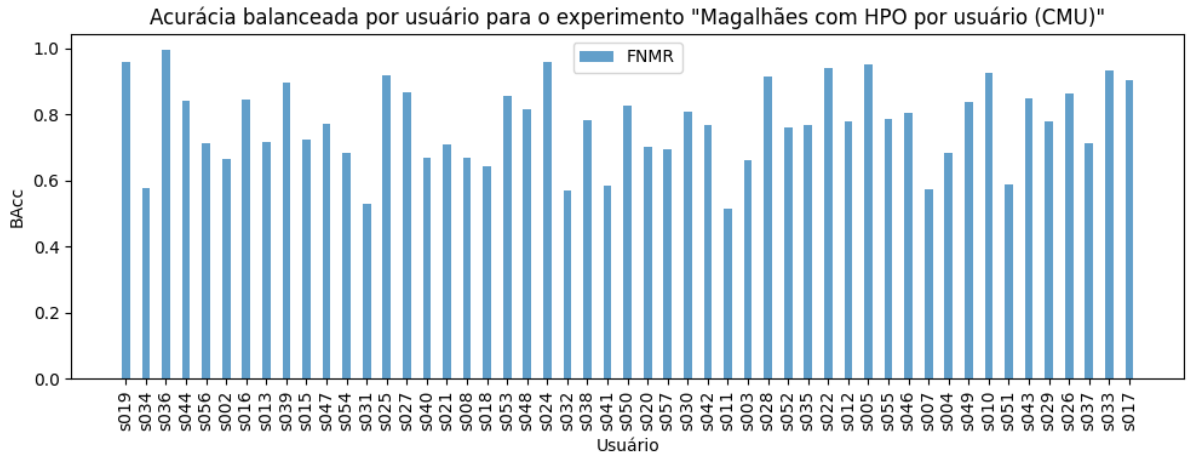


Figura 20 – Magalhães com HPO por usuário (Keyrecs) - BAcc

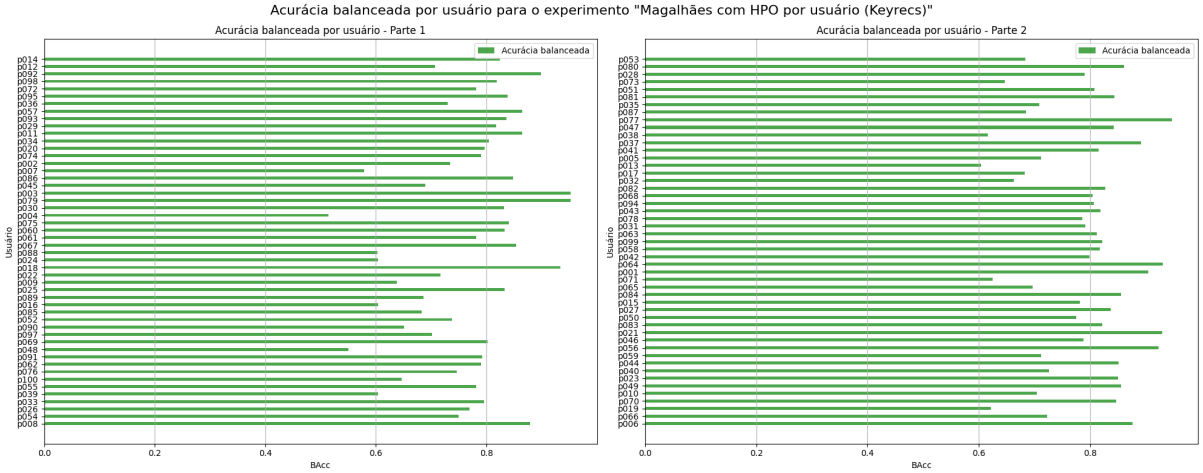


Figura 21 – Random Forest com HPO global (CMU) - BAcc

Figura 22 – Random Forest com HPO por usuário (CMU) - BAcc

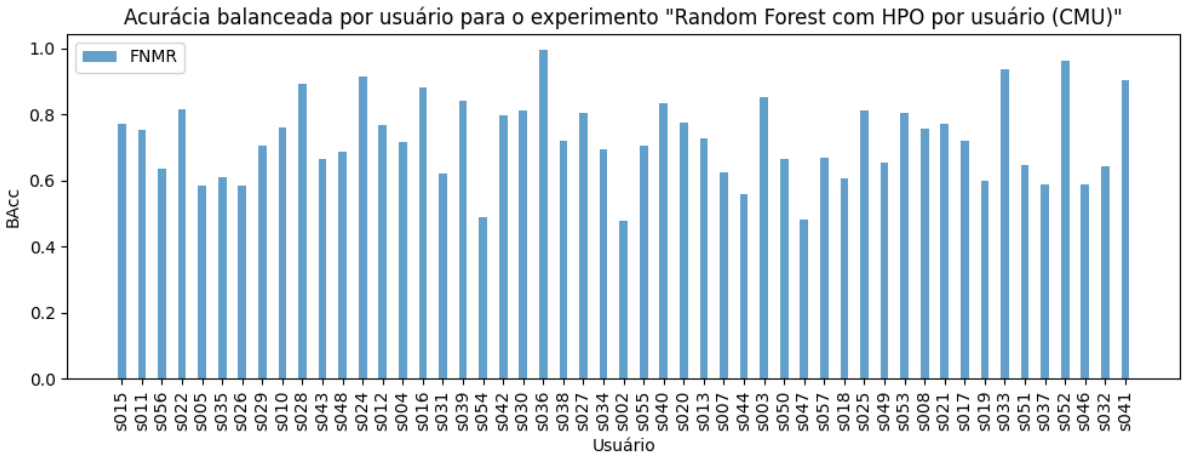


Figura 23 – SVM (CMU) - BAcc

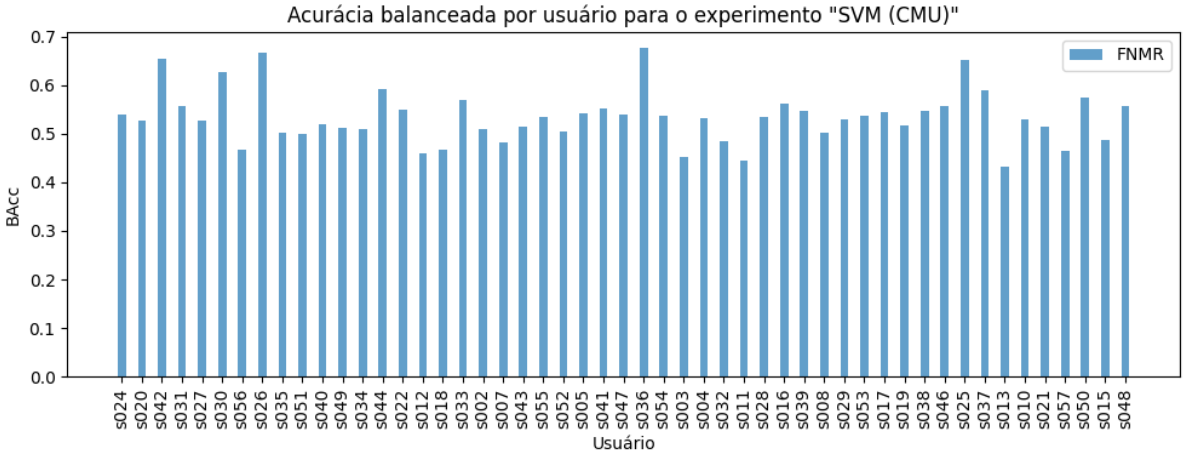


Figura 24 – SVM (Keyrecs) - BAcc

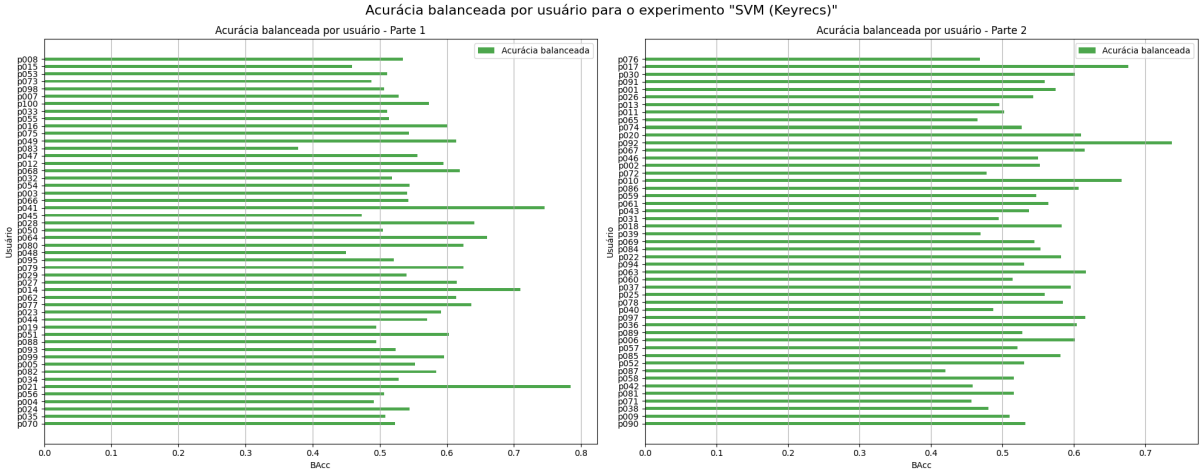


Figura 25 – SVM com HPO global (CMU) - BAcc

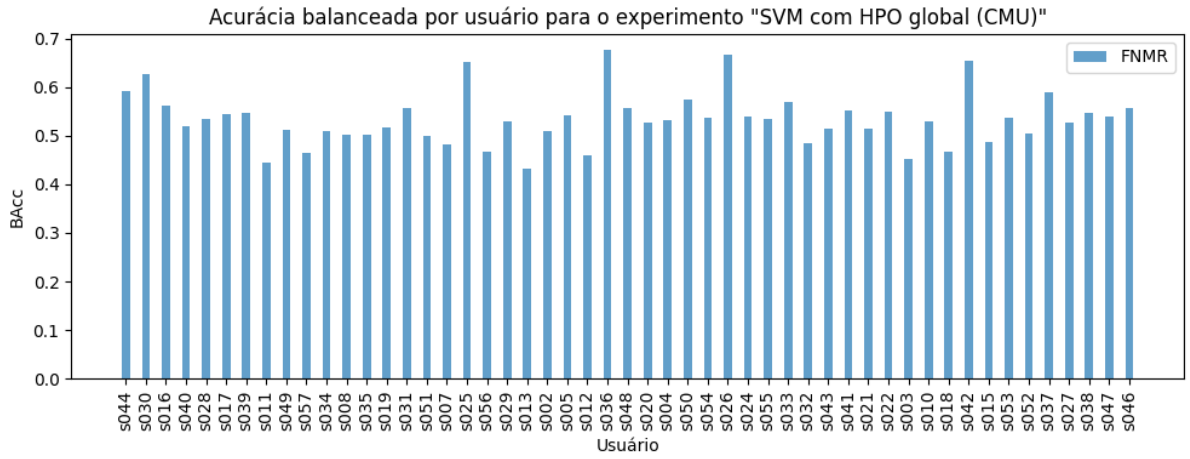


Figura 26 – SVM com HPO global (Keyrecs) - BAcc

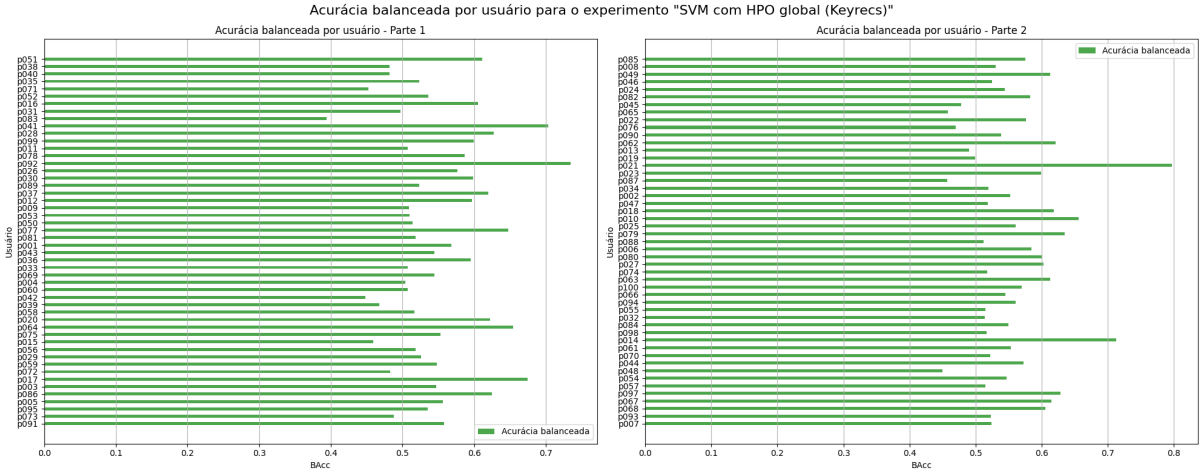


Figura 27 – SVM com HPO por usuário (CMU) - BAcc

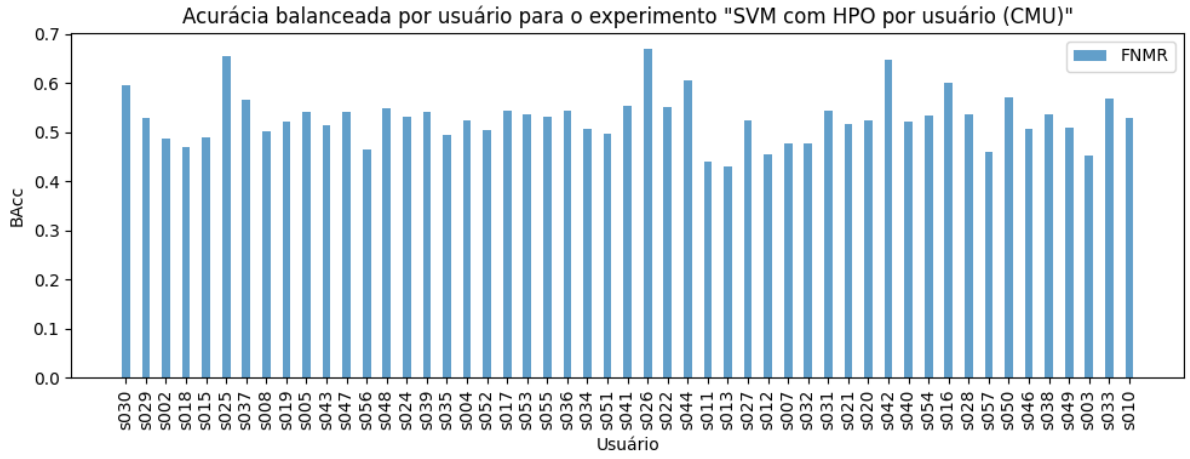
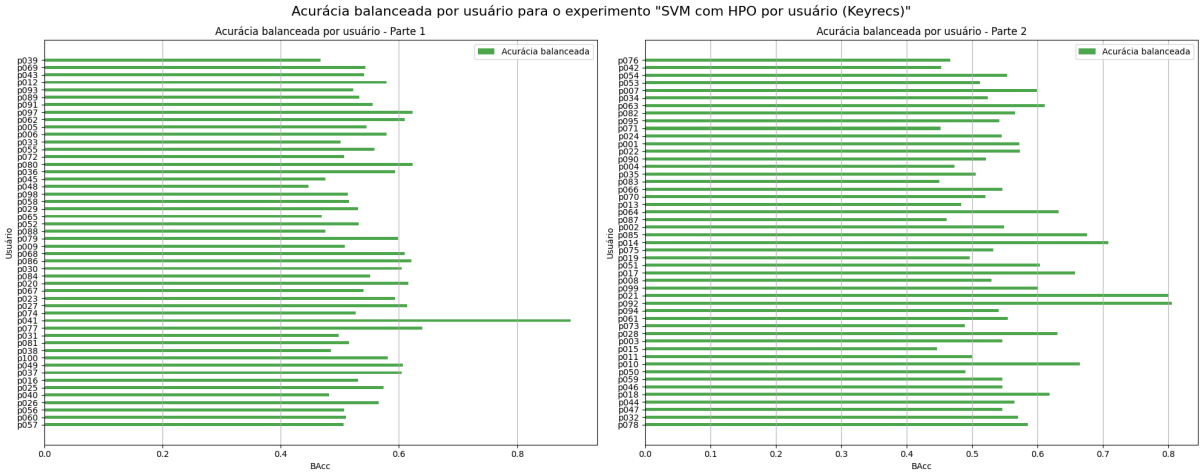


Figura 28 – SVM com HPO por usuário (Keyrecs) - BAcc



5 Conclusões e perspectivas de trabalhos futuros

Referências

BISCHL, B.; BINDER, M.; LANG, M.; PIELOK, T.; RICHTER, J.; COORS, S.; THOMAS, J.; ULLMANN, T.; BECKER, M.; BOULESTEIX, A.-L.; DENG, D.; LINDAUER, M. Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *WIREs Data Mining and Knowledge Discovery*, v. 13, n. 2, p. e1484, 2023.

DECASTRO-GARCÍA, N.; CASTAÑEDA, Á. L. M.; GARCÍA, D. E.; CARRIEGOS, M. V. Effect of the sampling of a dataset in the hyperparameter optimization phase over the efficiency of a machine learning algorithm. *Complexity*, Hindawi, v. 2019, p. 6278908, Feb 2019. ISSN 1076-2787.

DIAS, T.; VITORINO, J.; MAIA, E.; SOUSA, O.; PRAÇA, I. Keyrecs: A keystroke dynamics and typing pattern recognition dataset. *Data in Brief*, v. 50, p. 109509, 2023. ISSN 2352-3409.

FERLINI, A.; MA, D.; HARLE, R.; MASCOLO, C. Eargate: gait-based user identification with in-ear microphones. In: . New York, NY, USA: Association for Computing Machinery, 2021. (MobiCom '21), p. 337–349. ISBN 9781450383424.

GIOT, R.; EL-ABED, M.; HEMERY, B.; ROSENBERGER, C. Unconstrained keystroke dynamics authentication with shared secret. *Computers & Security*, v. 30, n. 6, p. 427 – 445, 2011. ISSN 0167-4048.

HUTTER, F.; KOTTHOFF, L.; VANSCHOREN, J. *Automated Machine Learning*. [S.l.]: Springer Cham, 2019. ISBN 978-3-030-05318-5.

JAIN, A.; FLYNN, P.; ROSS, A. *Handbook of Biometrics*. [S.l.]: Springer US, 2007. ISBN 9780387710419.

JAIN, A.; ROSS, A.; PRABHAKAR, S. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, IEEE, v. 14, n. 1, p. 4–20, 2004. ISSN 1051-8215.

JAIN, A. K.; NANDAKUMAR, K.; ROSS, A. 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recog. Letters*, Elsevier, v. 79, p. 80 – 105, 2016. ISSN 0167-8655.

KARNAN, M.; AKILA, M.; KRISHNARAJ, N. Biometric personal authentication using keystroke dynamics: A review. *Applied Soft Computing*, v. 11, n. 2, p. 1565–1573, 2011. ISSN 1568-4946. The Impact of Soft Computing for the Progress of Artificial Intelligence.

KASPROWSKI, P.; BOROWSKA, Z.; HAREZLAK, K. Biometric identification based on keystroke dynamics. *Sensors*, v. 22, n. 9, 2022. ISSN 1424-8220.

KILLOURHY, K. S.; MAXION, R. A. Comparing anomaly-detection algorithms for keystroke dynamics. *Proceedings of the International Conference on Dependable Systems and Networks*, p. 125–134, 2009. ISSN 9781424444212.

LU, X.; ZHANG, S.; HUI, P.; LIO, P. Continuous authentication by free-text keystroke based on cnn and rnn. *Computers & Security*, v. 96, p. 101861, 2020. ISSN 0167-4048.

MHENNI, A.; CHERRIER, E.; ROSENBERGER, C.; Essoukri Ben Amara, N. Double serial adaptation mechanism for keystroke dynamics authentication based on a single password. *Computers & Security*, v. 83, p. 151–166, 2019. ISSN 0167-4048. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167404818306059>.

MHENNI, A.; ROSENBERGER, C.; CHERRIER, E.; AMARA, N. E. B. Keystroke template update with adapted thresholds. In: . [S.l.]: Institute of Electrical and Electronics Engineers Inc., 2016. p. 483 – 488. ISBN 978-146738526-8.

MONROSE, F.; RUBIN, A. D. Keystroke dynamics as a biometric for authentication. *Future Generation Computer Systems*, v. 16, n. 4, p. 351 – 359, 2000. ISSN 0167-739X.

PEACOCK, A.; KE, X.; WILKERSON, M. Typing patterns: a key to user identification. *IEEE Security Privacy*, IEEE, v. 2, n. 5, p. 40–47, 2004.

Precise Biometrics. *Understanding Biometric Performance Evaluation*. 2014. Disponível em: <http://precisebiometrics.com/wp-content/uploads/2014/11/White-Paper-Understanding-Biometric-Performance-Evaluation.pdf>.

PURWAR, D. K.; VISHWAKARMA, D.; SINGH, N.; KHEMCHANDANI, V. One v/s all svm implementation for keystroke based authentication system. In: *2019 4th International Conference on Information Systems and Computer Networks (ISCON)*. [S.l.: s.n.], 2019. p. 268–272.

ROY, S.; PRADHAN, J.; KUMAR, A.; ADHIKARY, D. R. D.; ROY, U.; SINHA, D.; PAL, R. K. A systematic literature review on latest keystroke dynamics based models. *IEEE Access*, v. 10, p. 92192–92236, 2022.

RYU, R.; YEOM, S.; HERBERT, D.; DERMOUDY, J. The design and evaluation of adaptive biometric authentication systems: Current status, challenges and future direction. *ICT Express*, v. 9, n. 6, p. 1183–1197, 2023. ISSN 2405-9595.

SANTOS, H.; MAGALHãES, S. T. D.; SANTOS, H. M. D. *An Improved Statistical Keystroke Dynamics Algorithm*. [S.l.]. Disponível em: <https://www.researchgate.net/publication/52011524>.

YANG, L.; SHAMI, A. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, v. 415, p. 295–316, 2020. ISSN 0925-2312.