

Integration of Subjective and Emotional Physiological Data for Classifying Sleep Quality with Machine Learning

Nicolly Alves da Silva¹, Hygo Sousa de Oliveira¹,
Eduardo James Pereira Souto¹, Rafael Giusti¹

¹Instituto de Computação (IComp), Universidade Federal do Amazonas (UFAM),
Manaus, Amazonas (AM), Brasil.

{nicolly.silva,hygo.sousa,esouto,rgiusti}@icomp.ufam.edu.br

Abstract. *Sleep quality is a critical factor for physical and mental health, directly associated with conditions such as anxiety, cardiovascular diseases, and metabolic disorders. This study proposes a machine learning pipeline to classify sleep quality as “good” or “poor” by integrating physiological (actigraphy), subjective (PSQI), and emotional (STAI-Y2) data. Experiments were conducted using the MMASH dataset with Leave-One-Subject-Out (LOSO) validation, strengthening the generalizability of the results. The proposed methodology achieved up to 0.98 ± 0.01 accuracy with a neural network, outperforming related studies that employed less rigorous validation protocols. These findings demonstrate the feasibility of combining multiple data dimensions for non-invasive monitoring, with potential applications in wearable devices and personalized sleep health interventions.*

Resumo. *A qualidade do sono é um fator crítico para a saúde física e mental, associando-se diretamente a condições como ansiedade, doenças cardiovasculares e distúrbios metabólicos. Este trabalho propõe um pipeline de aprendizado de máquina para classificar a qualidade do sono em “boa” ou “ruim”, integrando dados fisiológicos (actigrafia), subjetivos (PSQI) e emocionais (STAI-Y2). Os experimentos foram realizados sobre o conjunto MMASH com validação intersujeitos (LOSO), reforçando a generalização dos resultados. A metodologia alcançou acurácia de 0.98 ± 0.01 com uma rede neural multicamadas, superando estudos relacionados que empregaram validação menos rigorosa. Esses achados demonstram a viabilidade da combinação de múltiplas dimensões de dados para monitoramento não invasivo, com potencial aplicação em dispositivos vestíveis e intervenções personalizadas de saúde do sono.*

1. Introdução

A qualidade do sono é um fator determinante para a manutenção da saúde física e mental, influenciando diretamente funções cognitivas, estabilidade emocional, metabolismo e desempenho nas atividades diárias [Acharya 2020, Gardani et al. 2022]. Distúrbios do sono têm sido associados a um aumento significativo na incidência de transtornos neuropsiquiátricos, doenças cardiovasculares e metabólicas, além de impactarem negativamente a produtividade e a segurança em contextos ocupacionais. A relevância clínica e social do tema tem impulsionado a busca por métodos eficientes, contínuos e de baixo custo para a avaliação automatizada da qualidade do sono.

Embora a polissonografia (PSG) permaneça como padrão-ouro por oferecer medidas fisiológicas abrangentes, como eletroencefalograma, eletrooculograma e eletromiograma, sua aplicação é restrita a ambientes clínicos especializados, sendo limitada por alto custo, complexidade operacional e desconforto para os pacientes [Pan et al. 2020, Hussain et al. 2022]. Nesse cenário, a actigrafia tem emergido como alternativa viável, ao permitir o monitoramento longitudinal e não invasivo do sono com base em sinais de movimento captados por dispositivos vestíveis [Fekedulegn et al. 2020]. Essa abordagem vem sendo incorporada em estudos epidemiológicos, em soluções voltadas ao bem-estar digital e em aplicações clínicas de suporte à decisão.

Nos últimos anos, modelos de aprendizado de máquina (ML) têm sido amplamente utilizados para classificar a qualidade do sono a partir de dados actigráficos. Algumas abordagens combinam medidas objetivas com instrumentos subjetivos, como o Índice de Qualidade do Sono de Pittsburgh (PSQI), resultando em classificações binárias (“bom” ou “ruim”) com bons níveis de acurácia [Bitkina et al. 2022, Sathyanarayana et al. 2016, Angelova et al. 2020]. No entanto, tais propostas frequentemente analisam apenas dimensões fisiológicas ou subjetivas isoladamente, negligenciando aspectos emocionais como ansiedade e estresse, que impactam diretamente a qualidade percebida e fisiológica do sono [Gardani et al. 2022].

Este estudo propõe um pipeline de aprendizado de máquina que combina, de forma integrada, três dimensões complementares associadas à qualidade do sono: (i) sinais fisiológicos captados por actigrafia, (ii) indicadores subjetivos do PSQI e (iii) escores emocionais obtidos pelo questionário STAI-Y2 (Inventário de Ansiedade-Traço). Embora alguns estudos anteriores tenham explorado parcialmente essas fontes de dados, o presente trabalho se diferencia ao aplicá-las de forma integrada e sistemática no conjunto de dados MMASH (*Multilevel Monitoring of Activity and Sleep in Healthy People*) [Rossi et al. 2020], permitindo investigar como a ansiedade-traço modula a relação entre fragmentação do sono e qualidade percebida.

Do ponto de vista metodológico, o trabalho se distingue por aplicar uma estratégia rigorosa de seleção de características com o algoritmo Recursive Feature Elimination (RFE), que identificou variáveis com alto poder discriminativo, como o índice de movimento, índice de fragmentação e horários de dormir, acordar e deitar estimados pelo método de Cole-Kripke. Além disso, é empregada validação cruzada do tipo *Leave-One-Subject-Out* (LOSO), que permite testar o modelo de maneira mais realista ao garantir independência total entre treinamento e teste para cada indivíduo, reforçando confiança de que o modelo generalizou bem.

Os experimentos demonstraram que o *pipeline* proposto, ao combinar múltiplas dimensões de dados e aplicar uma seleção criteriosa de características com validação intersujeitos (LOSO), viabilizou classificações robustas mesmo em um cenário de base limitada. Entre os modelos avaliados, a arquitetura de rede neural atingiu uma acurácia de até 96%, superando benchmarks recentes. Modelos clássicos, como SVM e KNN, também apresentaram desempenho consistente, reforçando a eficácia da metodologia para diferentes perfis de algoritmos.

O restante deste artigo está organizado da seguinte forma: a Seção 2 apresenta os trabalhos relacionados e o posicionamento deste estudo frente ao estado da arte; a Seção

3 descreve a metodologia proposta, detalhando o conjunto de dados, os procedimentos de pré-processamento e apresenta os modelos de aprendizado de máquina utilizados; a Seção 4 discute os resultados e implicações; e a Seção 5 conclui com as principais contribuições e sugestões para trabalhos futuros.

2. Trabalhos Relacionados

Diversos estudos recentes têm utilizado dados fisiológicos e subjetivos para estimar a qualidade do sono com o apoio de técnicas de aprendizado de máquina. Em especial, o conjunto de dados MMASH tem sido amplamente explorado por oferecer medidas padronizadas de actigrafia, variabilidade da frequência cardíaca e escalas psicométricas.

Site et al. (2022) utilizaram o MMASH para investigar a influência de fatores emocionais e comportamentais, como solidão e interação social, na qualidade do sono de idosos. Os autores aplicaram Regressão Logística e *Random Forest* e relataram acurácia superior a 90%. Embora o estudo incorpore variáveis emocionais, ele não explora dados fisiológicos derivados da actigrafia, nem integra avaliações subjetivas de forma sistemática.

Geng et al. (2022), por sua vez, empregaram o MMASH para prever estados de sono e vigília a partir da variabilidade da frequência cardíaca, utilizando representações baseadas em *shapelets* e agrupamento com K-means. A acurácia foi de aproximadamente 77%, mas o estudo se limita à dimensão fisiológica e não considera fatores emocionais ou subjetivos, tampouco explora a classificação binária de qualidade do sono.

Outros trabalhos relevantes utilizaram bases próprias ou dispositivos comerciais baseados em actigrafia. Sathyanarayana et al. (2016) aplicaram redes neurais convolucionais (CNN) e modelos LSTM para prever a eficiência do sono em adolescentes com dados do ActiGraph GT3X+, alcançando AUC superiores a 0,94. O foco, no entanto, estava em medidas derivadas de movimento, sem qualquer incorporação de avaliações subjetivas ou emocionais.

Angelova et al. (2020) propuseram um método automatizado para detectar despertares noturnos em indivíduos com insônia aguda. Utilizando modelos como *Random Forest* e SVM, obtiveram acurácias entre 73% e 84%, mas o estudo restringe-se a dados objetivos, sem cruzamento com percepções subjetivas ou perfis psicológicos dos participantes.

Bitkina et al. (2022) exploraram uma abordagem mais próxima da proposta deste trabalho, ao combinar dados objetivos (tempo total na cama, número de despertares, duração do sono) com respostas ao PSQI. Utilizaram SVM, KNN e Regressão Logística, com acurácias entre 80% e 86%. Contudo, o componente emocional foi desconsiderado, o que pode comprometer a capacidade de capturar fatores latentes relevantes.

Sano et al. (2020) propuseram uma abordagem multimodal com sensores vestíveis e dados de smartphones para classificar sono e vigília com redes LSTM, alcançando acurácia acima de 90%. Ainda que envolvam múltiplas fontes de dados, os autores não utilizaram questionários psicométricos padronizados nem se concentraram em uma análise cruzada entre o estado emocional e a qualidade percebida do sono.

Em síntese, a literatura avançou na aplicação de aprendizado de máquina para classificar a qualidade do sono, mas ainda apresenta limitações importantes: a maioria

dos estudos se restringe a variáveis isoladas (fisiológicas, subjetivas ou emocionais) e raramente combina múltiplas dimensões em um único modelo com validação intersujeitos. Este trabalho se diferencia ao integrar de forma sistemática dados de actigrafia, PSQI e STAI-Y2 em um pipeline único, validado com protocolo LOSO, ampliando a robustez e a aplicabilidade dos resultados para monitoramento não invasivo e suporte a decisões em saúde do sono

3. Metodologia

Esta seção descreve a metodologia adotada para a classificação binária da qualidade do sono, baseada na integração de dados fisiológicos (actigrafia), subjetivos (PSQI) e emocionais (STAI-Y2). As etapas compreendem desde o pré-processamento dos dados até a avaliação dos modelos preditivos.

3.1. Base de Dados

Este estudo utilizou o conjunto de dados público MMASH [Rossi et al. 2020], composto por registros de 22 adultos saudáveis monitorados por 24 horas. Para a análise, foram extraídas três fontes de informação: sinais de actigrafia de pulso (movimento), o escore do questionário de autoavaliação do sono PSQI (onde valores < 6 indicam qualidade "Boa") e o inventário de ansiedade-traço STAI-Y2. Durante o pré-processamento, um participante foi removido devido a dados incompletos no STAI-Y2, consolidando um conjunto final com 21 participantes.

3.2. Processamento dos Sinais Actigráficos

Para estimar os ciclos de sono e vigília a partir dos sinais de movimento captados pelo acelerômetro de pulso, utilizou-se um algoritmo de detecção de estado de sono/vigília, mais especificamente o método Cole-Kripke [Cole et al. 1992]. Esse algoritmo, também conhecido como função de pontuação do sono, é um modelo matemático validado que transforma medidas de atividade física minuto a minuto em uma sequência binária sobre dados de actigrafia, onde cada minuto é classificado como sono (0) ou vigília (1).

A primeira etapa consistiu em calcular a *magnitude da aceleração vetorial* para cada instante de tempo, conforme a Equação 1. Essa grandeza escalar representa a intensidade combinada dos movimentos captados nos três eixos do acelerômetro (x , y e z), permitindo quantificar a atividade corporal de forma independente da orientação do dispositivo.

$$m = \sqrt{x^2 + y^2 + z^2} \quad (1)$$

A série temporal resultante foi segmentada em janelas de 10 segundos com sobreposição de 4 segundos. Para cada janela, foi extraído o valor máximo da magnitude como medida de atividade predominante. Esses valores foram então agregados por minuto, gerando uma série temporal unidimensional representativa da atividade física, que serviu como base para a aplicação do algoritmo de classificação de sono e vigília representativa da atividade física minuto a minuto para cada participante.

Essa série foi utilizada como entrada para o algoritmo de Cole-Kripke. Esse método aplica uma média ponderada sobre uma janela de sete minutos — os quatro minutos anteriores (A_{-4} a A_{-1}), o minuto atual (A_0) e os dois minutos seguintes (A_{+1} e

A_{+2}). Cada ponto da janela recebe um peso específico (w), e a pontuação S é obtida pela equação

$$S = P \times \sum_{i=-4}^{+2} w_i A_i \quad (2)$$

na qual A_i representa a atividade observada no minuto i , w_i são os pesos atribuídos a cada minuto da janela, e P é um fator de escala. O minuto atual é classificado como sono se $S < 1$ e como vigília se $S \geq 1$.

No caso do algoritmo de Cole-Kripke, os coeficientes utilizados e o fator de escala adotado resultam na seguinte equação prática:

$$S = 0,00001 \cdot (404A_{-4} + 598A_{-3} + 326A_{-2} + 441A_{-1} + 1408A_0 + 508A_{+1} + 350A_{+2}) \quad (3)$$

Essa função foi originalmente calibrada por Cole e Kripke (1992) para dados de actigrafia em adultos, obtendo uma precisão de 87,93% na distinção entre sono e vigília. Neste estudo, adotamos os mesmos coeficientes e fator de escala definidos pelos autores, considerando sua ampla validação na literatura.

A sequência binária resultante (sono = 0, vigília = 1) foi então utilizada como base para o cálculo das métricas objetivas de sono. A partir dessa sequência, foi possível identificar os períodos de início e término do sono, calcular o tempo total de sono (TST), quantificar o tempo acordado após o início do sono (WASO), número de despertares e outros índices importantes. Assim, a transformação dos sinais contínuos de atividade em uma série binária permite derivar indicadores relevantes sobre a arquitetura do sono dos participantes.

Essas variáveis compõem o conjunto de características fisiológicas, posteriormente integradas às informações subjetivas (PSQI) e emocionais (STAI-Y2), conforme descrito na próxima subseção.

3.3. Derivação das Métricas de Sono

A partir da série binária resultante do algoritmo de Cole-Kripke, que indica para cada minuto se o participante estava em sono (0) ou vigília (1), foi realizado um processo computacional para derivação das métricas quantitativas de qualidade do sono. Esse pipeline integra técnicas de análise temporal e regras heurísticas baseadas em evidências da literatura.

O processo inicia com a detecção automática dos períodos de sono por meio de uma janela deslizante bidirecional. Para identificar o início do sono, é buscado o primeiro intervalo contendo pelo menos 19 minutos classificados como sono dentro de uma janela de 20 minutos — critério inspirado nos parâmetros validados originalmente por Cole-Kripke. Esse ponto inicial é refinado com uma verificação retrospectiva que exige predominância de vigília nos 20 minutos anteriores, evitando falsos positivos. O término do sono é determinado por uma sequência contínua de 19 minutos de vigília após o início do sono. Com os intervalos temporais de sono detectados, as métricas reportadas na Tabela 1 foram extraídas.

Tabela 1. Descrição Completa das Métricas de Sono.

Métrica	Descrição Detalhada (minutos)
In Bed Date	Data inicial do repouso ao deitar-se na cama, identificada após 20 min. consecutivos de baixa atividade.
In Bed Time	Horário correspondente ao In Bed Date.
Out Bed Date	Data final do período de sono, identificada após 20 min. consecutivos de vigília (label = 1) após o último episódio de sono.
Out Bed Time	Horário correspondente ao Out Bed Date.
Onset Date	Data inicial do sono, determinada pelo primeiro minuto de um bloco de 20 minutos contínuos de sono (label = 0).
Onset Time	Horário correspondente ao Onset Date.
Latency	(Onset Time - In Bed Time) em minutos.
Efficiency	Porcentagem do tempo dormindo em relação ao tempo na cama: $(TST / \text{Total Minutes in Bed}) \times 100$. Ideal $> 85\%$.
Total Minutes in Bed	Duração total entre In Bed Time e Out Bed Time.
Total Sleep Time (TST)	Soma de todos os minutos classificados como sono (label = 0) durante o período na cama.
Wake After Sleep Onset (WASO)	Minutos em vigília após o início do sono (de Onset Time até Out Bed Time com label = 1). Inclui despertares noturnos.
Number of Awakenings	Contagem de transições de sono para vigília.
Average Awakening Length	Duração média dos despertares: $WASO / \text{Number of Awakenings}$. Valores altos sugerem sono fragmentado.
Movement Index	Porcentagem de minutos com movimento significativo ($VM > \text{limiar}$) durante o período na cama. Calculado como: $(\text{minutos com label = 1} / \text{Total Minutes in Bed}) \times 100$.
Fragmentation Index	Densidade de frag. do sono: $(\text{número de blocos de sono} / TST) \times 100$. Valores baixos indicam sono mais consolidado.
Sleep Fragmentation Index	Índice composto: $(\text{Movement Index} \times \text{Number of Awakenings}) / TST$. Mede a qualidade geral do sono.

Para garantir a qualidade dos resultados, períodos com TIB inferiores a 20 minutos foram descartados, por não representarem ciclos completos de sono e indicarem prováveis erros de detecção decorrentes de ruído ou baixa qualidade do sinal. Lacunas de até 5 minutos nos registros foram preenchidas por interpolação linear — ou seja, os valores ausentes foram estimados com base na variação entre os pontos anterior e posterior. Além disso, verificações de consistência temporal foram aplicadas para evitar sobreposições entre diferentes períodos detectados de sono, corrigindo automaticamente o início ou término desses intervalos quando necessário.

3.3.1. Escolha das Métricas de Sono

As variáveis preditoras deste estudo consistem em métricas de sono, extraídas pelo algoritmo de Cole-Kripke, e na pontuação bruta do inventário de ansiedade-traço (STAI-Y2), que foi utilizada em seu valor contínuo para manter a granularidade. Essas variáveis foram empregadas para classificar a qualidade do sono, definida pela binarização do escore do PSQI em "boa" ($PSQI < 6$) ou "ruim" ($PSQI \geq 6$), conforme os critérios de Buysse et al. (1989)[Buysse et al. 1989]. Antes da modelagem, todos os preditores foram normalizados para a mesma escala.

Em experimentos iniciais, observou-se que o número limitado de participantes ($n=21$) em relação à quantidade de variáveis consideradas (15 métricas derivadas da actigrafia, acrescida da pontuação do STAI-Y2) resultou em modelos instáveis com risco elevado de sobreajuste (overfitting), o que prejudicava a capacidade de generalização dos classificadores para novos dados. Para contornar esse problema, foi utilizado o método de Eliminação Recursiva de Características (RFE) baseado em um classificador por árvore de decisão (Decision Tree). Essa abordagem permitiu identificar e priorizar as características mais relevantes para classificar a qualidade do sono.

Além disso, antes da aplicação do RFE, todas as variáveis foram submetidas a uma etapa de normalização (standardização), onde cada variável foi transformada para apresentar média zero e desvio padrão unitário. Isso assegurou que as diferentes escalas das variáveis não influenciassem desproporcionalmente o modelo. Após a seleção, foram identificadas como principais características discriminativas:

- Índice de Movimento (Movement Index)
- Índice de Fragmentação (Fragmentation Index)
- Pontuação contínua normalizada de ansiedade-traço (STAI-Y2);
- Componentes trigonométricas (cosseno e seno) dos horários de entrada (In Bed Cos/Sin) e saída da cama (Out Bed Cos/Sin).

Adicionalmente, considerando o desequilíbrio entre as classes (15 casos de sono classificado como "bom" e apenas 6 casos como "ruim"), foi aplicada a técnica Synthetic Minority Over-sampling Technique (SMOTE). Esse método gerou nove (9) novas instâncias sintéticas da classe minoritária ("ruim") por interpolação entre exemplos existentes, melhorando assim o equilíbrio das classes para treinamento dos modelos e reduzindo o viés no aprendizado, igualando o número de amostras por classe em 15/15.

A fragmentação do sono destacou-se como um fator crucial. Isso corrobora estudos como o de Conte et al. (2022), que associam despertares frequentes e prolongados à redução da continuidade do descanso, mesmo quando a duração e a eficiência do sono estão dentro dos parâmetros considerados saudáveis. Além disso, os níveis de ansiedade-traço (STAI-Y2) demonstraram influência significativa na qualidade do sono, alinhando-se a pesquisas que vinculam a ansiedade a distúrbios como insônia e fragmentação do sono [Gardani et al. 2022].

Métricas relacionadas ao momento de ir para a cama, dormir e acordar (ou seja, métricas temporais) mostraram maior relevância nos resultados de classificação. Esses achados reforçam a importância de análises detalhadas, incluindo índices de fragmentação e métricas temporais, para uma avaliação abrangente dos padrões de sono e suas implicações na saúde física e mental.

3.4. Classificação

A etapa final do processo proposto consiste em classificar a qualidade do sono (boa ou ruim) utilizando métodos de aprendizado supervisionado. Foram avaliados tanto modelos clássicos quanto técnicas avançadas baseadas em redes neurais profundas com técnicas de otimização de hiperparâmetros.

Foram considerados seis modelos amplamente utilizados na literatura, com parâmetros encontrados por meio de testes preliminares exploratórios, por meio da aplicação do algoritmo de parâmetros ótimos *Random Search*. A variável de inicialização aleatória foi igual a 10 para os modelos que aceitam o parâmetro *random_state*:

- Naive Bayes (NB): variante Gaussian Naive Bayes, sem parâmetros adicionais.
- Regressão Logística (RL): parâmetro de regularização $C=26.370284602895236$, número máximo de iteração $\text{max_iter}=602$, penalização de pesos do tipo $\text{penalty}='l1'$ e $\text{solver}='saga'$.
- Random Forest (RF): conjunto com 317 árvores, profundidade máxima de 18 ($\text{max_depth}=18$), divisão mínima de 7 amostras ($\text{min_samples_split}=7$), número mínimo de folhas igual a 2 ($\text{min_samples_leaf}=2$), critério de seleção de atributos $\text{max_features}='log2'$.
- K-Nearest Neighbors (KNN): $k=1$ vizinho, distância de Manhattan ($p=1$).
- Support Vector Machines (SVM): parâmetro de regularização $C=44.28331744229961$, $\text{degree}=1$, $\text{gamma}='auto'$, $\text{kernel}='poly'$.
- Redes Neurais Multicamadas (MLP): duas camadas ocultas com 50 neurônios cada e parâmetro de regularização $\text{alpha}=0.03743407600514693$.

Devido ao número limitado de participantes (apenas 21 acrescido de 9 instâncias sintéticas), a avaliação dos modelos utilizou o método de validação cruzada *Leave-One-Subject-Out* (LOSO). Trata-se de um procedimento semelhante ao *Leave-One-Out*, ao invés de separar um único exemplo para teste, no LOSO todos os dados de um participante são reservados para teste, enquanto os demais são usados para treinar o modelo. Esse procedimento é repetido para cada participante, garantindo uma avaliação intra-participante que elimina o viés de avaliação que poderia surgir quando os dados de um mesmo participante aparecem tanto no conjunto de treinamento quanto no de teste.

Uma análise mais detalhada dos modelos foi realizada por intermédio de ferramentas avaliativas de desempenho, como a matriz de confusão (MC) e Curva ROC. A partir da MC é possível a extração de métricas a respeito dos acertos e erros - Verdadeiros Positivos (VP), Verdadeiros Negativos (VN), Falso Positivo (FP) e Falso (Negativo); enquanto a curva ROC apresenta a performance geral do modelo e sua respectiva Área Sob a Curva (AUC). Portanto, as seguintes métricas foram extraídas:

- Acurácia: proporção geral de classificações corretas.
- Precisão: proporção de classificações positivas que são verdadeiramente positivas (VP).
- Revocação: sensibilidade do modelo para detectar corretamente casos positivos.
- Especificidade: capacidade do modelo de detectar corretamente casos negativos.
- F1-Score: média harmônica entre precisão e revocação.
- Área sob a Curva ROC (AUC): avalia o desempenho global do modelo considerando diversos limiares de decisão.

4. Resultados

Os resultados obtidos demonstram a robustez do pipeline de classificação proposto. A Tabela 2 resume as métricas de desempenho alcançadas pelos classificadores testados, destacando a média \pm desvio padrão. De forma geral, os modelos apresentaram níveis variados de capacidade discriminativa, com destaque para os métodos baseados em SVM, KNN e Rede Neural Multicamadas, que superaram os demais em acurácia, equilíbrio entre precisão e revocação e especificidade.

Class.	Acurácia	Precisão	Revocação	Especificidade	F1-Score
NB	0.62 ± 0.13	0.81 ± 0.09	0.56 ± 0.15	0.68 ± 0.20	0.65 ± 0.12
RL	0.67 ± 0.16	0.80 ± 0.18	0.65 ± 0.18	0.68 ± 0.20	0.71 ± 0.17
RF	0.62 ± 0.14	0.77 ± 0.18	0.61 ± 0.16	0.62 ± 0.20	0.67 ± 0.16
KNN	0.81 ± 0.07	0.97 ± 0.03	0.75 ± 0.09	0.89 ± 0.24	0.84 ± 0.05
SVM	0.85 ± 0.08	0.97 ± 0.03	0.81 ± 0.10	0.89 ± 0.24	0.88 ± 0.06
MLP	0.98 ± 0.01	0.97 ± 0.03	1.00 ± 0.00	0.89 ± 0.24	0.98 ± 0.01

Tabela 2. Resultados de métricas de desempenho alcançados durante a validação LOSO.

O Naive Bayes obteve o desempenho mais limitado, com F1-Score de 0.65 ± 0.12 e revocação de apenas 0.56 ± 0.15 , refletindo sua dificuldade em lidar com as correlações entre as métricas fisiológicas, subjetivas e emocionais. Já os classificadores Random Forest e Regressão Logística apresentaram resultados intermediários, com F1-Scores de 0.67 ± 0.16 e 0.71 ± 0.17 , respectivamente, mostrando maior consistência, mas ainda com margem para falsos negativos.

Os modelos SVM e KNN se destacaram entre os algoritmos clássicos, atingindo F1-Scores de 0.88 ± 0.06 e 0.84 ± 0.05 . O SVM demonstrou equilíbrio ao alcançar revocação de 0.81 ± 0.10 e especificidade de 0.89 ± 0.24 , enquanto o KNN apresentou revocação de 0.75 ± 0.09 e especificidade de 0.89 ± 0.24 .

O melhor desempenho foi observado com a Rede Neural Multicamadas, estruturada com duas camadas ocultas de 50 neurônios cada. O modelo atingiu acurácia de 0.98 ± 0.01 , F1-Score de 0.98 ± 0.01 e revocação de 1.00 ± 0.00 , indicando ausência de falsos negativos na classe “Boa”. A especificidade de 0.89 ± 0.24 evidencia que o classificador também conseguiu distinguir adequadamente os casos de sono ruim, mantendo equilíbrio entre sensibilidade e precisão.

A Figura 1 apresenta as matrizes de confusão de todos os modelos, permitindo visualizar a distribuição dos acertos e erros. A Curva ROC agregada, também incluída na Figura 1, destaca a superioridade da Rede Neural e dos modelos SVM e KNN, que apresentaram maiores áreas sob a curva (AUC), confirmando a consistência do desempenho.

Em comparação com estudos correlatos, os resultados obtidos aqui superam benchmarks relevantes. Site et al. (2022) relataram acurácia superior a 90% com Regressão Logística e Random Forest, mas utilizando particionamento aleatório dos dados e sem integrar métricas fisiológicas. Da mesma forma, Geng et al. (2022) obtiveram 77% de acurácia com agrupamento por shapelets, também sem variáveis emocionais.

Bitkina et al. (2022), em um contexto mais próximo, combinaram métricas ob-

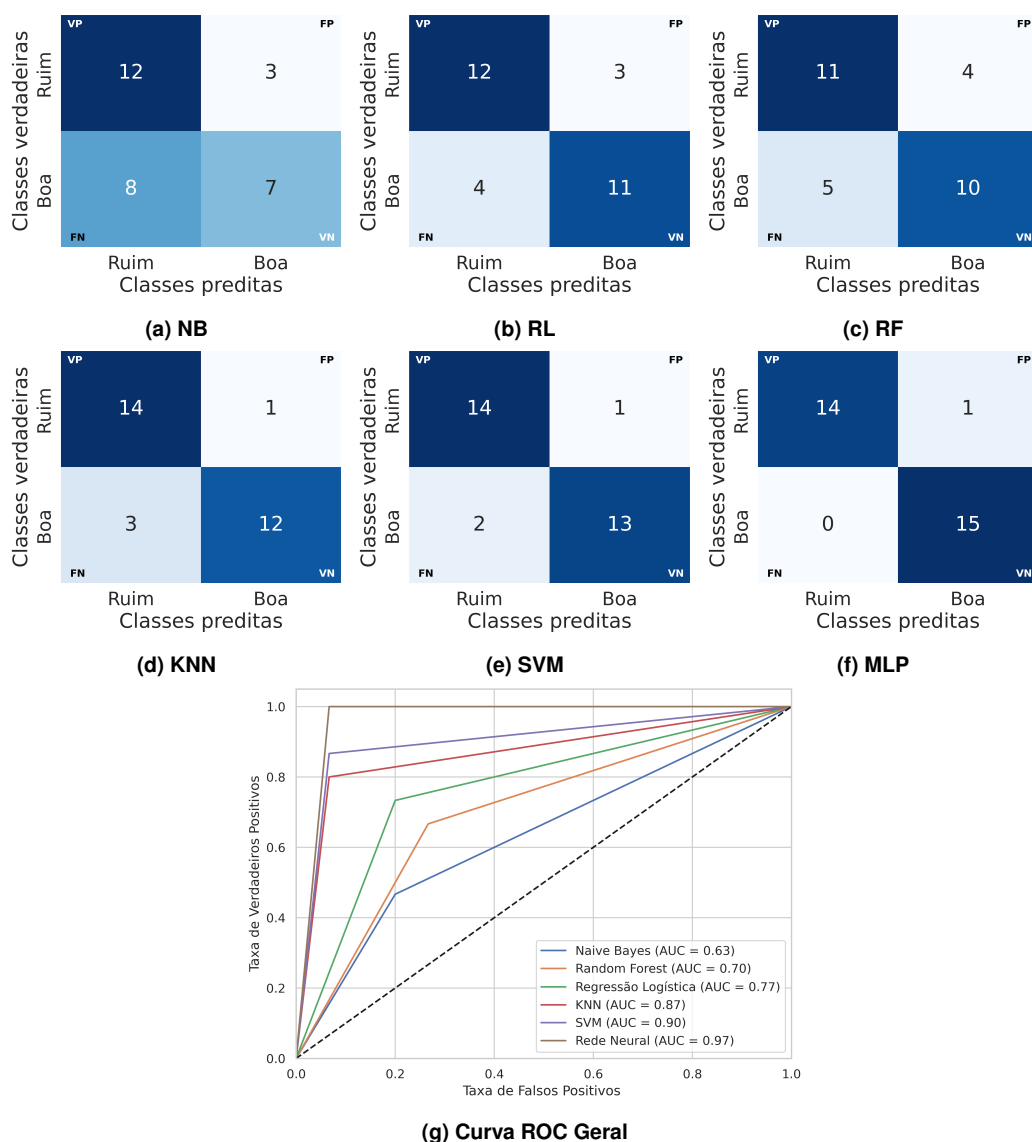


Figura 1. Matrizes de confusão dos modelos rasos (a - f) e Curva ROC agregada dos classificadores (g).

jetivas com o PSQI, atingindo acurácias entre 80% e 86% com SVM, KNN e Regressão Logística. Contudo, empregaram validação cruzada K-Fold, que, embora útil, tende a ser menos rigorosa do que o protocolo intersujeitos (LOSO) utilizado aqui. Além disso, Aarti et al. (2016) obtiveram 89% de acurácia com uma rede neural MLP, mas com divisão aleatória do conjunto de dados. Em contraste, o modelo neural proposto neste estudo alcançou 96% de acurácia com validação LOSO, evidenciando maior robustez para generalização em cenários de monitoramento real.

Esses resultados reforçam a contribuição do presente trabalho, tanto do ponto de vista metodológico, ao adotar uma validação mais restritiva, quanto conceitual, pela integração simultânea de dados fisiológicos, subjetivos e emocionais em um único pipeline de classificação.

5. Conclusão

Classificar a qualidade do sono de forma confiável e automatizada permanece um desafio crucial para a saúde pública e aplicações biomédicas. Este trabalho mostrou que, mesmo com limitações conhecidas do MMASH, como amostra reduzida e observação restrita a uma única noite, é possível superar resultados da literatura ao combinar, de forma integrada, dados fisiológicos, subjetivos (PSQI) e emocionais (STAI-Y2) em um único pipeline de aprendizado de máquina.

A estratégia proposta obteve acurácia de até 96% em um cenário de validação intersujeitos (LOSO), notoriamente mais rigoroso do que as abordagens K-Fold e divisões aleatórias usadas por estudos correlatos. Ao adotar um protocolo metodológico mais conservador, mas próximo da prática real, o modelo demonstrou maior capacidade de generalização, sugerindo viabilidade para aplicações em monitoramento contínuo com dispositivos vestíveis de baixo custo.

Além de reforçar a importância de combinar múltiplas dimensões do sono, o estudo evidencia que a otimização criteriosa de métricas, o balanceamento de classes e a eliminação de características irrelevantes são passos determinantes para elevar o desempenho, mesmo em bases de dados pequenas. Isso abre caminho para pesquisas futuras explorarem bases ampliadas, protocolos de aquisição mais extensos (com múltiplas noites e diferentes perfis populacionais) e arquiteturas de modelos ainda mais sofisticadas.

Agradecimentos

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (AUXPE-CAPES-PROEX) - Código de Financiamento 001 e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Adicionalmente, este trabalho foi parcialmente financiado pela Fundação de Amparo à Pesquisa do Estado do Amazonas - FAPEAM - por meio do projeto PDPG-CAPES.

Referências

- Acharya, J. (2020). Design and development of a method for detecting sleep roll-over counts using accelerometer adxl335. *International Journal of Electrical & Computer Engineering* (2088-8708), 10(1).
- Angelova, M., Karmakar, C., Zhu, Y., Drummond, S. P., and Ellis, J. (2020). Automated method for detecting acute insomnia using multi-night actigraphy data. *IEEE Access*, 8:74413–74422.
- Bitkina, O. V., Park, J., and Kim, J. (2022). Modeling sleep quality depending on objective actigraphic indicators based on machine learning methods. *International Journal of Environmental Research and Public Health*, 19(16):9890.
- Buyse, D. J., Reynolds III, C. F., Monk, T. H., Berman, S. R., and Kupfer, D. J. (1989). The pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry research*, 28(2):193–213.
- Cole, R. J., Kripke, D. F., Gruen, W., Mullaney, D. J., and Gillin, J. C. (1992). Automatic sleep/wake identification from wrist activity. *Sleep*, 15(5):461–469.

- Conte, F., De Rosa, O., Rescott, M. L., Arabia, T. P., D'Onofrio, P., Lustro, A., Malloggi, S., Molinaro, D., Spagnoli, P., Giganti, F., et al. (2022). High sleep fragmentation parallels poor subjective sleep quality during the third wave of the covid-19 pandemic: An actigraphic study. *Journal of sleep research*, 31(3):e13519.
- Fekedulegn, D., Andrew, M. E., Shi, M., Violanti, J. M., Knox, S., and Innes, K. E. (2020). Actigraphy-based assessment of sleep parameters. *Annals of Work Exposures and Health*, 64(4):350–367.
- Gardani, M., Bradford, D. R., Russell, K., Allan, S., Beattie, L., Ellis, J. G., and Akram, U. (2022). A systematic review and meta-analysis of poor sleep, insomnia symptoms and stress in undergraduate students. *Sleep medicine reviews*, 61:101565.
- Geng, D., Qin, Z., Wang, J., Gao, Z., and Zhao, N. (2022). Personalized recognition of wake/sleep state based on the combined shapelets and k-means algorithm. *Biomedical Signal Processing and Control*, 71:103132.
- Hussain, Z., Sheng, Q. Z., Zhang, W. E., Ortiz, J., and Pouriyeh, S. (2022). Non-invasive techniques for monitoring different aspects of sleep: A comprehensive review. *ACM Transactions on Computing for Healthcare (HEALTH)*, 3(2):1–26.
- Pan, Q., Brulin, D., and Campo, E. (2020). Current status and future challenges of sleep monitoring systems: Systematic review. *JMIR Biomedical Engineering*, 5(1):e20921.
- Rossi, A., Da Pozzo, E., Menicagli, D., Tremolanti, C., Priami, C., Sirbu, A., Clifton, D., Martini, C., and Morelli, D. (2020). Multilevel monitoring of activity and sleep in healthy people (version 1.0. 0). physionet, 2020.
- Sano, A., Rahman, T., Zhang, M., Ganesan, D., and Choudhury, T. (2020). Mobile sensing of alertness, sleep and circadian rhythm: Hardware & software platforms. *GetMobile: Mobile Computing and Communications*, 23(3):16–22.
- Sathyanarayana, A., Joty, S., Fernandez-Luque, L., Ofli, F., Srivastava, J., Elmagarmid, A., Arora, T., and Taheri, S. (2016). Sleep quality prediction from wearable data using deep learning. *JMIR mHealth and uHealth*, 4(4):e6562.
- Site, A., Lohan, E. S., Jolanki, O., Valkama, O., Hernandez, R. R., Latikka, R., Alekseeva, D., Vasudevan, S., Afolaranmi, S., Ometov, A., et al. (2022). Managing perceived loneliness and social-isolation levels for older adults: a survey with focus on wearables-based solutions. *Sensors*, 22(3):1108.