

**MINISTERUL EDUCAȚIEI ȘI CERCETĂRII AL REPUBLICII
MOLDOVA**

**Universitatea Tehnică a Moldovei
Facultatea Calculatoare, Informatică și Microelectronică
Departamentul Microelectronică și Inginerie Biomedicală**

**Admis la susținere
Șef interimar DMIB:
Railean Serghei, conf. univ., dr.**

« ____ » _____ 2025

**APLICAREA TEHNOLOGIILOR DE ÎNVĂȚARE
AUTOMATĂ DESTINAT ANALIZEI
SEMNALELOR EEG**

Teză de master

Student:

**Topal Artiom,
IBM-231M**

Coordonator:

**Pocaznoi Ion,
conf. univ., dr.**

Chișinău, 2025

REZUMAT

la teza de master cu tema «**Aplicarea tehnologiilor de învățare automată destinat analizei semnalelor EEG**»,

Teza cuprinde fundamentele teoretice despre activitatea bioelectrică a creierului și utilizarea electroencefalografiei (EEG) în scopuri diagnostice, precum și o analiză a metodelor cheie de învățare automată. Lucrarea se concentrează în detaliu pe utilizarea metodei cu vectori de suport (SVM) pentru detectarea automată a epilepsiei. Partea practică descrie mediul de dezvoltare, preprocesarea și filtrarea înregistrărilor EEG reale, extragerea caracteristicilor informative și ajustarea hiperparametrilor SVM folosind Python. Sunt prezentate rezultatele experimentale, incluzând acuratețea, matricea de confuzie și o vizualizare a hiperplanei de decizie în spațiul bidimensional al componentelor principale.

Cuvinte cheie: electroencefalografie, epilepsie, metodă cu vectori de suport, învățare automată, preprocesare semnal, caracteristici wavelet.

Scopul lucrării este de a elabora și valida un algoritm de învățare automată capabil să clasifice în mod fiabil segmente scurte de EEG ca fiind patologice (epilepsie) sau normale.

Obiectivele generale au constat în selectarea și configurarea mediului de dezvoltare (Python, scikit-learn, MNE), unificarea și filtrarea înregistrărilor EEG, extragerea automată a caracteristicilor (temporale, spectrale și bazate pe wavelet) și reducerea dimensiunii (PCA), precum și analiza comparativă a rezultatelor pentru diferite parametri SVM.

Metodele aplicate au urmat abordarea clasică de învățare supravegheată: împărțirea datelor în seturi de antrenament și test cu validare încrucișată, ajustarea hiperparametrilor SVM (C și γ) și vizualizarea rezultatelor într-un spațiu bidimensional al componentelor principale.

Rezultatele obținute indică faptul că metoda cu vectori de suport cu nucleu radial și hiperparametri optimizați poate atinge o precizie de aproximativ 95–97% în diferențierea semnalelor „epilepsie vs. normal” chiar și în condiții de date limitate. Experimentele au arătat că o preprocesare atentă (îndepărtarea artefactelor, alinierea canalelor, filtrarea semnalelor) și o selecție judicioasă a caracteristicilor sunt factori critici pentru performanța finală a algoritmului.

Valoarea aplicativă a lucrării: constă în posibilitatea de a integra sistemul propus ca modul automatizat de susținere a diagnosticului de epilepsie în mediul clinic. Rezultatele confirmă viabilitatea aplicării învățării automate în analiza biosemnalelor, cu potențial de extindere la alte afecțiuni neurologice.

ANNOTATION

to the master's thesis addressing the topic «**Application of machine learning technologies for analyzing EEG signals**»,

This thesis provides a concise theoretical rationale of the brain's bioelectrical activity and the use of electroencephalography (EEG) in diagnostics, reviews key machine learning methods, and offers a detailed examination of how Support Vector Machines (SVM) can be employed for automated epilepsy detection. The practical part describes the development environment, preprocessing and filtering of real EEG records, the extraction of informative features, and the tuning of SVM hyperparameters using Python. Experimental results are presented, including accuracy assessment, confusion matrices, and a visualization of the decision boundary in a two-dimensional principal component space.

Keywords: electroencephalography, epilepsy, support vector machine, machine learning, signal preprocessing, wavelet features.

The aim of this work is to develop and validate a machine learning algorithm capable of reliably classifying short EEG segments as either pathological (epilepsy) or healthy.

General tasks included choosing and configuring a development environment (Python, scikit-learn, MNE), unifying and filtering the EEG recordings, automatically extracting features (temporal, spectral, wavelet-based) and reducing dimensionality (PCA), as well as conducting a comparative analysis of results under different SVM parameters.

Methods used followed the classic supervised learning approach: splitting data into training and test sets with cross-validation, tuning of SVM hyperparameters (C and γ), and visualizing the results in a two-dimensional principal component space.

Results obtained show that a support vector machine with a radial basis kernel and optimized hyperparameters can achieve around 95–97% accuracy in distinguishing “epilepsy vs. normal” signals under limited data conditions. Experiments revealed that thorough preprocessing (artifact removal, channel alignment, signal filtering) and careful feature selection are critical for final algorithm performance.

The practical significance of this work lies in demonstrating how the proposed system can form the basis of an automated diagnostic support module for epilepsy in healthcare settings. The results highlight the feasibility of machine learning for analyzing real biosignals, with potential extension to other neurological pathologies.

АННОТАЦИЯ

к магистерской дипломной работе на тему «**Применение технологий машинного обучения для анализа сигналов ЭЭГ**»,

Дипломная работа включает краткое теоретическое обоснование биоэлектрической активности мозга и применения электроэнцефалографии (ЭЭГ) в диагностике, обзор ключевых методов машинного обучения и детальный разбор использования метода опорных векторов (SVM) для автоматизированного выявления эпилепсии. В практической части описана среда разработки, предобработка и фильтрация реальных ЭЭГ-записей, извлечение информативных признаков и настройка гиперпараметров SVM с помощью средств Python. Приведены результаты экспериментов с оценкой точности, матрицами ошибок и визуализацией разделяющей гиперплоскости в двумерном пространстве главных компонент.

Ключевые слова: электроэнцефалография, эпилепсия, метод опорных векторов, машинное обучение, предобработка сигналов, вейвлет-признаки.

Целью данной работы является разработка и проверка алгоритма машинного обучения, способного надежно классифицировать короткие фрагменты ЭЭГ как патологические (эпилепсия) или здоровые.

Общие задачи заключались в выборе и настройке среды разработки (Python, scikit-learn, MNE), в унификации и фильтрации ЭЭГ-записей, в автоматизированном извлечении признаков (временные, спектральные, вейвлет-признаки) и снижении размерности (PCA), а также в сравнительном анализе результатов при разных параметрах SVM.

Применяемые методы включали классическую схему обучения с учителем: разбиение на обучающую и тестовую выборки с кросс-валидацией, подбор гиперпараметров SVM (C и γ) и визуализацию результатов в двумерном пространстве главных компонент.

Полученные результаты демонстрируют, что метод опорных векторов, использующий радиально-базисное ядро и оптимизированные гиперпараметры, способен достичь точности около 95–97% в задаче классификации «эпилепсия / норма» при ограниченных данных.

Прикладное значение работы состоит в том, что предложенная система может послужить основой для автоматизированного модуля поддержки диагностики эпилепсии в медицинских учреждениях. Результаты демонстрируют жизнеспособность машинного обучения при анализе реальных биосигналов при неврологических патологиях.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	8
1 ОСНОВЫ ЭЛЕКТРОЭНЦЕФАЛОГРАФИИ	12
1.1 Нейрофизиологические основы биоэлектрической активности мозга.....	12
1.2 Методика регистрации сигнала и запись электроэнцефалограммы.....	15
1.3 Ритмы электроэнцефалограммы.....	17
1.4 Диагностика патологий при помощи электроэнцефалограммы.....	18
2 ОСНОВЫ МАШИННОГО ОБУЧЕНИЯ	24
2.1 Определение и классификация методов машинного обучения.....	24
2.1.1 Супервизорное машинное обучение.....	25
2.1.2 Несупервизорное машинное обучение.....	28
2.1.3 Машинное обучение с подкреплением.....	31
2.2 Применимость машинного обучения в анализе ЭЭГ.....	32
2.3 Метод опорных векторов.....	38
3 АНАЛИЗ ЭЭГ МЕТОДОМ ОПОРНЫХ ВЕКТОРОВ	45
3.1 Среда разработки и инструментарий.....	49
3.2 Сбор и предобработка ЭЭГ-данных.....	52
3.3 Выделение и отбор признаков.....	55
3.4 Настройка и обучение модели.....	58
3.5 Результаты и оценка точности.....	61
ЗАКЛЮЧЕНИЕ	67
БИБЛИОГРАФИЯ	70
ПРИЛОЖЕНИЯ	72
Приложение 1. Листинг кода Microsoft VSC: анализ ЭЭГ методом опорных векторов.....	72

ВВЕДЕНИЕ

В последние десятилетия стремительное развитие технологий машинного обучения и искусственного интеллекта открывает новые перспективы в различных областях науки и техники. В частности, одной из таких областей является биомедицинская инженерия, находящаяся на стыке таких научных дисциплин, как медицина, информатика, математика, физика и иных. Так, медицина предоставляет клинический контекст и понимание заболеваний, информатика и машинное обучение предлагают методы обработки и анализа сложных биомедицинских данных, а математика обеспечивает теоретическую основу для разработки и оптимизации алгоритмов. Подобный междисциплинарный подход позволяет интегрировать знания и инструменты из различных областей для решения сложных задач, стоящих перед современной медициной. В то же время, одной из первостепенных целей биомедицинской инженерии является стремление к существенному и качественному улучшению методов своевременной диагностики и лечения заболеваний.

Опираясь на вышеобозначенные тезисы, представленное в данной дипломной работе исследование ставит перед собой задачу осуществить разработку соответствующего метода анализа результатов записи биоэлектрической активности мозга (электроэнцефалограмм) с целью их последующей классификации методами машинного обучения на предмет здоровой либо же имеющей признаки тех или иных неврологических заболеваний мозговой активности. Иными словами, имея достаточный объем данных записанных энцефалограмм здоровых и нездоровых с неврологической точки зрения пациентов, обучить на этих данных методами машинного обучения модель с целью в последующем получить возможность программно прогнозировать наличие/отсутствие патологии на новых данных и тем самым обеспечить техническое содействие врачебному персоналу в принятии решения о постановке диагноза.

Актуальность и значимость исследования в области разработки программно-технических методов исследования биоэлектрической активности мозга подтверждаются социальными и статистическими исследованиями. Так, согласно данным Всемирной Организации Здравоохранения [14], по состоянию на 2021 год более трех миллиардов человек в мире страдали неврологическими заболеваниями, которые, в свою очередь, являются ведущей причиной нездоровья и инвалидности. Помимо уже хорошо известных ведущих неврологических причин утраты здоровья, таких как инсульты, энцефалопатия, эпилепсия и нейродегенеративные заболевания, опасную тенденцию к распространению проявляют и ранее не существовавшие неврологические расстройства, вызванные

осложнениями инфекции COVID-19 (например, в виде нарушений когнитивной функции и синдрома Гийена – Барре).

В то же время, электроэнцефалография (ЭЭГ), как неинвазивный метод исследования и оценки биоэлектрической активности мозга, остается наиболее широко применяемым методом в неврологической практике. Об этом свидетельствует как клиническая статистика, так и актуальные научные исследования в этой области, нацеленные на модернизацию метода. Электроэнцефалография является одним из ключевых методов исследования функциональной активности головного мозга и позволяет выявлять характерные паттерны электрической активности, ассоциированные с различными неврологическими состояниями. Так, согласно усредненным данным [3], соответствие заболевания с процентом его успешного обнаружения одним лишь ЭЭГ-исследованием следующее: для коматозных состояний и оценки глубины наркоза ЭЭГ успешен более чем на 90%; для нарушений сна (нарколепсия) - до 90%; для эпилепсии – от 80% до 90%; для цереброваскулярных заболеваний (инсульты) - около 80%; для нейродегенеративных заболеваний - от 60% до 70%.

Суммируя вышеизложенное и ориентируясь на поставленную задачу исследования, дипломная работа имеет, за исключением данной вводной главы, следующую структуру. В первой главе было подробно рассмотрено, что такое электроэнцефалография, как мозг порождает биоэлектрические сигналы и как эти сигналы регистрируются с помощью электроэнцефалографа. Была показана клиническая ценность метода и раскрыта его зависимость от ритмической структуры мозга: выделяли альфа-, бета-, дельта-, тета-ритмы, указывали их типичные частотные диапазоны и клиническую интерпретацию. Эта теоретическая основа стала необходимым фундаментом, позволившем более детально понять, какими характеристиками сигнала следует оперировать в задачах автоматизированной диагностики. Вопросы, посвящённые диагностике патологий по ЭЭГ, позволили уяснить, что существует целый спектр неврологических состояний, для которых данный метод остаётся одним из базовых инструментов выявления расстройств. Однако на практике врачу приходится тратить значительное время на тщательный просмотр и анализ данных, а тонкие паттерны могут остаться незамеченными из-за человеческого фактора или больших объёмов исследования. Всё это предопределяет важность развития автоматических методов анализа, использующих передовые технологии машинного обучения.

Вторая глава была отведена обзору машинного обучения как научного направления, где был дан обзор базовой терминологии и классификации методов. Особое внимание

уделялось тому, почему эти методы так широко применяются именно в биомедицине и конкретно в анализе ЭЭГ. Акцент был сделан на двух ключевых моментах: необходимость работать с большим массивом цифровых сигналов и сложность их интерпретации в случае, если паттерны слабо выражены или глубоко спрятаны за шумом и артефактами. В результате выделили различные парадигмы машинного обучения, среди которых для данной работы оказались наиболее интересны методы супервизорного (контролируемого) обучения. Конкретные примеры применения в медицине показали, что классификация эпилепсии, поиск ранних маркеров деменции, оценка стадий сна — всё это становится более доступным, если удастся составить качественные признаки из ЭЭГ-сигнала. В этом же разделе подробно разобран метод опорных векторов: его математическая основа, возможность гибкого выбора нелинейного ядра, а также механизм мягкого зазора, позволяющий учитывать шум и небольшой процент ошибок на обучающих данных. Была подчеркнута высокая эффективность метода при работе с биомедицинскими датасетами, если имеется небольшое число записей, но четко определённые классы.

На этой теоретической базе в третьей главе перешли к детальному описанию собственно выполненной практической части, где основной задачей было продемонстрировать, как именно организовать анализ ЭЭГ методом опорных векторов, начиная с подготовки данных и заканчивая финальной визуализацией результатов. Сначала в разделе о среде разработки было показано, что для этой цели хорошо подходит язык Python с его многочисленными библиотеками (mne, scikit-learn, NumPy, pandas и прочие), а в качестве интегрированной среды разработки можно использовать Microsoft Visual Studio. Поясняется, почему удобна модульная структура кода и как она способствует воспроизводимости исследования.

Затем в разделе по сбору и предобработке ЭЭГ-данных приводились конкретные операции по загрузке материалов из открытой платформы, переименованию каналов, учёту артефактов и формированию 60-секундных фрагментов, пригодных для сравнения между пациентами. Ключевым моментом здесь оказалось раздельное отношение к «здоровым» испытуемым и к пациентам с эпилепсией, включая создание канала A2-A1 у последних. Далее подробно освещались вопросы фильтрации сигнала, приведения к единому набору каналов и оценка достаточной длительности записи, чтобы гарантировать минимум шума и артефактов.

Вслед за этим шла подглава, посвящённая выделению и отбору признаков. На основе временных сегментов (эпох) применялись методы вейвлет-преобразования (в частности, «mexh» вейвлет) и вычислялись статистические показатели (среднее, дисперсия,

асимметрия). Эти признаки затем агрегировались по эпохам и каналам, образуя единый вектор для каждого субъекта. Дополнительная нормализация и метод главных компонент снижали размерность и устраняли мультиколлинеарность, способствуя более надёжной работе метода опорных векторов.

Далее в разделе «Настройка и обучение модели» обсуждалась основная часть процесса машинного обучения. Сформированные признаковые вектора и их метки подавались в метод опорных векторов, который, используя радиально-базисное ядро, тренировался на перекрёстной проверке Leave-One-Out. Была подчеркнута логика выбора гиперпараметров модели и то, как мягкий зазор и нелинейное ядро позволяют учесть разнообразие сигналов. Проводились эксперименты с различными настройками, и, согласно итогам, перебор гиперпараметров выявлял «сильные» комбинации, повышающие точность классификации и особенно чувствительность к пациентам с эпилепсией.

Наконец, в последнем параграфе «Результаты и оценка точности» приводились количественные итоги работы. Изначально модель с параметрами по умолчанию давала точность около 89–90%, но вызывалось смещение в пользу здоровых и недодиагностирование части эпилептических случаев. Оптимизация гиперпараметров увеличила итоговую точность модели (свыше 95–96%), сократив процент неверных решений. Подробно рассматривались метрики, показывающие сбалансированное распознавание обоих классов, а также матрица ошибок, наглядно иллюстрирующая, как часто модель путает здоровых и больных. Ещё один аспект заключался в обсуждении вычислительных затрат модели и алгоритма машинного обучения.

В итоге, анализ проведённого исследования подтверждает гипотезу о том, что метод опорных векторов, особенно при корректном выборе гиперпараметров и тщательной предобработке сигналов (удалении артефактов, выделении информативных признаков), способен достигать высоких показателей точности при решении задачи классификации «эпилепсия / норма» на основе коротких фрагментов ЭЭГ. Данные результаты согласуются с многочисленными работами в сфере биомедицинской информатики, где метод опорных векторов исторически зарекомендовал себя как устойчивый, хорошо интерпретируемый и достаточно гибкий алгоритм. Таким образом, можно заключить, что выбранная стратегия оказалась результативной, а возможности SVM и современных инструментов машинного обучения позволяют провести успешную автоматизированную диагностику эпилепсии по данным ЭЭГ даже при относительно небольшом размере выборки и умеренных вычислительных ресурсах.

БИБЛИОГРАФИЯ

1. ЗЕНКОВ, Л. Р. *Клиническая электроэнцефалография (с элементами эпилептологии). Руководство для врачей.* Москва: МЕДпресс-информ, 2018. 360 p. ISBN 978-5-00030-583-6.
2. NIEDERMEYER, Ernst, LOPES DA SILVA, Fernando. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields.* Philadelphia: Lippincott Williams&Wilkins, 2005. 1309 p. ISBN 978-0781751261.
3. BLINOWSKA, K., DURKA, P. Electroencephalography (EEG). In: *Wiley Encyclopedia of Biomedical Engineering.* Hoboken: John Wiley&Sons, 2006, pp. 249-295. ISBN 978-0-471-24967-2.
4. ВЬЮГИН, В. В. *Математические основы теории машинного обучения и прогнозирования.* Москва: МЦМНО, 2013. 387 p. ISBN 978-5-4439-2014-6.
5. БУРКОВ, Андрей. *Машинное обучение без лишних слов.* Санкт-Петербург: Питер, 2020. 192 p. ISBN 978-5-4461-1560-0.
6. ГРИГОРЬЕВ, Алексей. *Машинное обучение. Портфолио реальных проектов.* Санкт-Петербург: Питер, 2023. 496 p. ISBN 978-5-4461-1978-3.
7. BRINK, H., RICHARDS, J.W., FETHEROLF, M. *Real-World Machine Learning.* New York City: Manning Publications, 2016. 264 p. ISBN 978-1617291920.
8. HOSSEINI, M-P., HOSSEINI, A., AHI, K. A Review on Machine Learning for EEG Signal Processing in Bioengineering. In: *IEEE Reviews in Biomedical Engineering.* 2021, vol 14, pp. 204-218. ISSN 1941-1189.
9. SAEIDI, M., KARWOWSKI, W., FARAHANI, F.V., FIOK, K., TAIAR, R., HANCOCK, P.A., AL-JUAID, A. Neural Decoding of EEG Signals with Machine Learning: A Systematic Review. In: *Brain Sciences.* 2021, nr 11(11), pp. 1525-1569. ISSN 2076-3425.
10. SACCA, V., CAMPOLO, M., MIRARCHI, D., GAMBARDELLA, A., VELTRI, P., MORABITO, F.P. On the Classification of EEG Signal by Using an SVM Based Algorithm. In: *Multidisciplinary Approaches to Neural Computing.* Berlin: Springer, 2017, pp. 271-278. ISBN 9783319569031.
11. WISE, A., SHANMUGADASS, V., SURESH, G.R. EEG Signal Analysis for Automated Epilepsy Seizure Detection Using Wavelet Transform and Artificial Neural Network. In: *Journal of Medical Imaging and Health Informatics.* 2019, nr 9(6), pp. 1301-1306. ISSN 2156-7018.

12. *OpenNeuro*. Stanford Center for Reproducible Neuroscience, © 2025 [citat 5.01.2025].
Disponibil: <https://openneuro.org>
13. *Biroul national de statistică al Republicii Moldova*. Biroul national de statistică al Republicii Moldova, © 2025 [citat 4.01.2024]. Disponibil: <https://statistica.gov.md/ro/>
14. *Health data overview for the Republic of Moldova*. World Health Organization, © 2024 [citat 3.01.2025]. Disponibil: <https://data.who.int/countries/498>
15. MICROSOFT. Visual Studio Code 1.93.1 [software]. 13 September 2024. [citat 20.10.2024]. Disponibil: <https://visualstudio.microsoft.com/>