

## ABSTRACT

**of the dissertation work by Ikram Medeuuly Bazarbekov titled “Research of artificial intelligence models and methods for diagnosing diseases with impaired cognitive functions”, submitted for the degree of Doctor of Philosophy (PhD) in the educational program 8D06102 – «Computer and Software Engineering»**

### Relevance

Diseases accompanied by impairment of cognitive processes represent an extensive class of neurodegenerative disorders, which are one of the main medical and social problems of modernity. Within this class of pathologies, Alzheimer’s disease (AD) occupies a dominant position, accounting for up to 70% of all clinical cases of dementia. In this regard, in the present dissertation work, Alzheimer’s disease was chosen as a base model for the study and development of diagnostic methods, the results of which can subsequently be extrapolated to other types of cognitive disorders.

AD is the most common cause of dementia, affecting millions of people worldwide. According to WHO (World Health Organization) forecasts, the number of patients will triple by 2050. Existing diagnostic methods, such as MRI (magnetic resonance imaging), PET (positron emission tomography), CT (computed tomography), EEG (electroencephalography), MEG (magnetoencephalography) and cerebrospinal fluid analysis provide high accuracy, but are expensive, invasive, or poorly accessible for mass screening. Standard cognitive tests, such as MMSE (Mini-mental state examination), MoCA (Montreal cognitive assessment), are often insufficiently sensitive at early stages. In this regard, the need to search for new digital biomarkers is increasing.

It is known that motor control, namely fine motor skills of writing, begins to deteriorate long before the manifestation of obvious cognitive symptoms. The application of artificial intelligence methods to the analysis of handwriting kinematics opens new possibilities for creating objective and scalable diagnostic tools. However, the introduction of deep learning into clinical practice is often limited by the lack of labeled datasets. Thus, the development of methods that allow effective training of artificial intelligence (AI) models on limited samples, including using synthetic data, is an extremely urgent task.

Thus, the relevance of this study is обусловлена by the growth in the number of patients with AD and the limitations of existing screening tools. Standard imaging methods, such as PET, require the use of radioactive tracers and have limited accessibility, while MRI may be contraindicated in patients with metal implants or claustrophobia. In addition, there are no standardized mechanisms for continuous dynamic monitoring of disease progression. Therefore, there is an urgent need for cost-effective, non-invasive and objective digital biomarkers that can be implemented at the primary level of healthcare to identify individuals at risk before significant neurological damage occurs.

The main **goal** of this work is the study and evaluation of methods for early diagnosis of AD based on artificial intelligence. It предполагает the integration of

new sources of sensor data, in particular handwriting kinematics, with complex deep learning architectures and data augmentation methods to improve the accuracy, sensitivity and clinical accessibility of diagnosing this disease.

To achieve the stated goal, it was necessary to solve the following **tasks**:

1) To conduct a critical review and generalization of modern AI and machine learning (ML) methods used in various instrumental research methods, including MRI, PET, EEG, MEG and sensor data;

2) To develop the architecture and prototype of a sensor device for recording kinematic characteristics of graphomotor activity reflecting motor control disorders;

3) To form an experimental sample by collecting handwriting kinematic parameters in clinical conditions from representative groups of subjects, including patients diagnosed with Alzheimer's disease and healthy volunteers;

4) To investigate and quantitatively evaluate specific kinematic differences in handwriting (for example, speed, acceleration and tremor) between patients with AD and cognitively healthy people from the control group;

5) To implement and evaluate a method for generating synthetic data based on mathematical modeling of pathological patterns to solve the problem of limited clinical datasets by generating synthetic data based on physical models and simulating AD symptoms;

6) To conduct a comparative study of neural network architectures based on recurrent and transformer models (LSTM, LSTM with attention mechanism, Transformer Encoder), as well as to design a hybrid deep learning architecture CNN-BiLSTM and determine the best classification quality among all studied approaches in the task of detecting cognitive decline at an early stage.

The **object** of the research is the process of diagnosis and monitoring of neurodegenerative diseases associated with cognitive impairments.

The **subject** of the research is artificial intelligence models, methods of generating synthetic data for augmentation of training samples, as well as kinematic digital biomarkers obtained using sensor writing instruments used for cognitive assessment.

### **Methodology**

The methodological basis of the work is the collection and analysis of kinematic data obtained from inertial sensors (IMU) during the performance of writing tasks. For signal preprocessing and calculation of biomechanical parameters, methods of digital filtering and mathematical modeling were used. The problem of limited sample size was solved by stochastic generation of synthetic data. The main classification tool is the developed hybrid neural network architecture CNN-BiLSTM, which allows identifying spatiotemporal patterns of pathology. Model verification was carried out using cross-validation with evaluation of accuracy, sensitivity and area under the ROC curve.

### **Scientific provisions submitted for defense**

1) The architecture and technical implementation of a sensor instrument (Smart Pen) based on an inertial module, providing high-frequency registration (100 Hz) of hand movement kinematics in 6 degrees of freedom, which makes it possible

to objectify hidden disorders of fine motor skills that are inaccessible to standard assessment methods.

2) The method of forming an 18-channel feature space based on biomechanical time series of IMU signals, including derivative and integral kinematic characteristics (jerk, velocity, angular accelerations, orientation), the statistical analysis of which confirms the diagnostic significance of the identified intergroup differences between patients with Alzheimer's disease and healthy subjects ( $p < 0.05$ ).

3) A method for generating synthetic data based on mathematical modeling of pathological patterns, including algorithms for stochastic noise injection (to simulate tremor) and nonlinear temporal deformation (to simulate bradykinesia), the application of which makes it possible to compensate for the deficit of clinical samples and prevent model overfitting.

4) A hybrid architecture of a deep neural network (CNN-BiLSTM), implementing the synergy of convolutional layers for detecting local anomalies (microtremor) and bidirectional recurrent layers for analyzing temporal dynamics (cognitive pauses), which ensures predictive accuracy AUC 0.96, significantly exceeding classical machine learning algorithms.

#### **Main results of the research**

1) A comparative analysis of modern methods for diagnosing neurodegenerative diseases was carried out, which revealed significant limitations of current approaches (high cost of neuroimaging, subjectivity of clinical scales) and substantiated the need to develop accessible instrumental screening tools based on objective analysis of fine motor skills.

2) A sensor device was developed for recording handwriting kinematic parameters, including a combined inertial sensor (accelerometer, gyroscope) and specialized software, which makes it possible to conduct non-invasive data collection in natural writing conditions, eliminating stress effects on the patient typical for laboratory tests.

3) An experimental database of 3D handwriting kinematics was collected, characterized by high resolution and inclusion of dynamic parameters (jerk, microtremor), which ensures the necessary representativeness of the sample for identifying hidden patterns of neurodegeneration and verifying deep learning models.

4) A data augmentation method was applied for training neural networks under small sample conditions, based on mathematical modeling of pathological motor patterns (tremor and bradykinesia) using stochastic algorithms and nonlinear time scale deformation, which makes it possible to compensate for the lack of clinical examples and increase the generalization ability of models.

5) A comparative study of neural network architectures based on recurrent and transformer models was conducted: basic LSTM achieved AUC 0.7779, LSTM+Attention model – AUC 0.9597, Transformer Encoder – AUC 0.9268; a hybrid CNN-BiLSTM architecture was designed and validated, which demonstrated the best results among all studied approaches (AUC 0.963, Recall 0.930), which

confirms the superiority of hybrid architectures over both classical machine learning algorithms and recurrent neural networks.

6) A set of kinematic indicators was identified, consisting of jerk, tremor and acceleration parameters, which allows diagnosing early stages of cognitive impairment with a sensitivity of 93.4 percent, outperforming standard clinical tests in effectiveness.

### **Scientific novelty**

The scientific novelty of the research lies in the development of a hybrid neural network architecture (CNN-BiLSTM) for analyzing dynamic time series of handwriting kinematics, as well as a method for generating synthetic samples based on mathematical modeling of pathological symptoms (tremor and bradykinesia), which supplemented for the first time a uniquely formed experimental database of labeled fine motor signals in network training. This approach makes it possible to compensate for the lack of clinical data and achieve high predictive accuracy (AUC 0.96) compared to classical machine learning methods.

### **Theoretical significance**

The theoretical significance of the dissertation work lies in the scientific substantiation of the effectiveness of applying deep learning methods for diagnosing neurodegenerative diseases under conditions of limited data samples. The work proves that the method of combined training, based on the joint use of real clinical records and synthetic data, makes it possible to compensate for the lack of medical statistics. It is established that mathematical modeling of the physical nature of symptoms, namely tremor and bradykinesia, provides the necessary representativeness of the training sample, which guarantees high accuracy and stability of predictive models.

A significant contribution to theory is the substantiation of the expediency of using a hybrid neural network architecture for analyzing handwriting kinematics. It is established that combining convolutional layers for detecting local features and recurrent layers for analyzing temporal dynamics makes it possible to identify hidden motor disorders that cannot be detected by single-component methods.

### **Practical significance**

The practical significance of the dissertation work lies in the development and implementation of applied tools for automated diagnosis of cognitive impairments.

Firstly, a sensor device based on inertial measurement modules was developed. This device is a cost-effective and mobile solution that allows non-invasive mass screening of the population at the primary healthcare level without requiring special technical training of medical personnel.

Secondly, deep learning models implementing the proposed algorithms for processing and classifying biomechanical time series of IMU signals were developed and experimentally validated. The effectiveness was confirmed experimentally: a pathology recognition accuracy of 93.4% was achieved, which exceeds the performance of traditional machine learning methods.

Thirdly, the formed annotated database of three-dimensional handwriting kinematic parameters is of practical value. The created dataset is a unique resource

for verifying diagnostic models and can be used by research groups for developing new methods for analyzing neurodegenerative diseases.

### **Reliability of results**

The reliability of the obtained results and the validity of the scientific provisions of the dissertation are confirmed by the correct application of the fundamental mathematical apparatus, including methods of mathematical statistics, digital signal processing, and deep learning. Of particular importance is the use of physically grounded models of motion kinematics, such as stochastic modeling of tremor and nonlinear deformation of the time scale, which ensures the adequacy of synthetic data generation algorithms to real biomechanical processes.

The empirical basis of the study was a verified dataset obtained during clinical experiments with the participation of a representative sample of 215 subjects, including 106 patients with a confirmed diagnosis of Alzheimer's disease and 109 healthy volunteers. The volume of the formed sample is sufficient to obtain statistically significant estimates with the required confidence probability. High accuracy of recording primary physical parameters is guaranteed by hardware verification of the developed Smart Pen device based on the calibrated inertial module MPU-9250, which ensures reproducibility of experiments.

Evaluation of the effectiveness of the developed neural network models was carried out using generally accepted standardized metrics in international practice, including sensitivity, specificity, and the area under the ROC curve. To eliminate the effect of overfitting and confirm the stability of the results, a k-fold cross-validation procedure was applied. Evidence of reliability is also provided by the conducted comparative analysis, which experimentally confirmed the advantage of the proposed hybrid approach over classical machine learning algorithms and single-component neural networks on identical test samples.

### **Approbation of the dissertation work**

The main provisions and practical results of the dissertation work were presented, discussed, and received a positive assessment at international scientific and technical conferences.

The results of this study were also presented in the following publications:

1) Bazarbekov, I., Almisreb, A., Ipalakova, M., Bazarbekova, M., Daineko, Y. (2026). Sim-to-Real Domain Adaptation for Early Alzheimer's Detection from Handwriting Kinematics Using Hybrid Deep Learning. *Sensors*, 26(1), 298. <https://doi.org/10.3390/s26010298>;

2) Bazarbekov, I., Razaque, A., Ipalakova, M., Yoo, J., Assipova, Z., & Almisreb, A. (2024). A review of artificial intelligence methods for Alzheimer's disease diagnosis: Insights from neuroimaging to sensor data analysis. *Biomedical Signal Processing and Control*, 92, 106023. <https://doi.org/10.1016/j.bspc.2024.106023>;

3) Базарбеков И.М., Ипалакова М.Т., Дайнеко Е.А., Муханов С.Б. Альцгеймер ауруын диагностикалауға арналған робo-пен-нің деректерін өңдеу және талдау: алдын ала нәтижелер. Қазақ-Британ Техникалық Университетінің хабаршысы. 2024; 21(3):78-89. <https://doi.org/10.55452/1998-6688-2024-21-3-78-89>;

4) Bazarbekov, I.; Ipalakova, M.; Daineko, Y.; Mukhanov, S.; Bazarbekova, M.; Zholdassova, Z.; Turgunova, A.; Kapyshev, G. Design of a smart handwriting tool for early detection of Alzheimer’s disease. In Proceedings of the 2025 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA), Antalya, Turkiye, 7–9 August 2025; IEEE: Piscataway, NJ, USA, 2025; pp. 1–6.

5) Yegemberdiyev, T., Daineko, Y., & Bazarbekov, I. (2025). Efficiency of artificial intelligence in the diagnosis of cognitive disorders. *Procedia Computer Science*, 265, 620–624. <https://doi.org/10.1016/j.procs.2025.07.229>;

6) Bazarbekov I.M., Ipalakova M.T., Daineko E.A., Zholdassova Z.A., Bazarbekova M.M., Assipova Zh.M. Certificate of state registration of a scientific work No. 56208 dated 31.03.2025 of the Republic of Kazakhstan. Methodology for collecting and describing handwriting motor data for cognitive impairment analysis.

In all listed publications, the applicant has a leading role in problem formulation, conducting research, and preparing the main text of the articles.

### **Connection with state programs**

The dissertation work was carried out within the framework of grant funding for young scientists under the project “Zhas Galym” (IRN AP25796437) on the topic “Development of an intelligent system for diagnosing cognitive impairments based on machine learning methods and handwriting motor analysis”.

The research corresponds to the strategic priorities of the development of the Republic of Kazakhstan and contributes to the implementation of the Concept for the Development of Artificial Intelligence in the Republic of Kazakhstan for 2024–2029 in terms of the development and implementation of domestic intelligent systems for supporting medical decision-making and the use of machine learning technologies for analyzing biomedical data. The developed hybrid neural network model and data augmentation method contribute to the development of national potential in the field of AI.

### **Main content of the dissertation**

The dissertation work consists of an introduction, four sections, a conclusion, a list of references, and appendices. The logic of the study reflects a последовательный переход from the analysis of existing scientific approaches to the development of an own engineering-algorithmic system and experimental verification of the proposed solutions.

In the first section, a theoretical analysis of clinical aspects of neurodegenerative diseases is presented using Alzheimer’s disease as an example. Modern instrumental diagnostic methods are considered, including neuroimaging and neurophysiological technologies, and a review of machine learning methods used in medical diagnostics is carried out. Special attention is paid to the analysis of handwriting kinematic parameters as a promising digital biomarker of early cognitive decline. The section ends with the formulation of the scientific problem and justification of the need to develop a new approach.

In the second section, the engineering implementation of a sensor system for recording handwriting kinematic parameters is described. The justification for choosing the hardware platform, device architecture, and principles of integration of

inertial modules is presented. The stages of design, prototyping, and software implementation of the data acquisition and transmission module are detailed. The methodology of conducting the experiment and the characteristics of the formed clinical sample are also given.

In the third section, algorithms for digital signal processing, extraction of diagnostically significant features, and formation of training datasets are developed. The results of a comparative analysis of handwriting kinematic characteristics in patient groups and the control sample are presented. A method for generating synthetic data based on mathematical modeling of pathological motor patterns has been developed, aimed at compensating for the limited clinical sample and improving the generalization ability of models.

In the fourth section, classical and deep machine learning methods are implemented and experimentally investigated. Tuning and comparative analysis of SVM, Random Forest, k-NN, and logistic regression algorithms are performed. A comparative study of neural network architectures based on recurrent and transformer models is conducted: basic LSTM, LSTM with soft attention mechanism, and Transformer Encoder. A hybrid CNN-BiLSTM architecture for analyzing spatiotemporal patterns of biomechanical time series of IMU signals is substantiated and developed, which demonstrated the best results among all studied approaches. Hyperparameter optimization and evaluation of model performance using standardized quality metrics are carried out.

In the conclusion, the main results of the study are formulated, scientific and practical conclusions are summarized, and перспективные направления дальнейших исследований in the field of applying artificial intelligence methods for diagnosing cognitive impairments are determined.

The dissertation contains 101 pages including 21 figures, 12 tables, and 2 appendices. References consist of 132 manuscripts.