Report

Thesis: “Early Detection of Breast Cancer Using Machine Learning techniques”.

Authors: Ibrayeva M., Ibrayeva T.
Scientific supervisor: Nurtas M.

Relevance. Breast cancer detection using machine learning techniques is an active area of research and development. In recent years, there have been significant advancements in the use of machine learning algorithms to analyze large datasets of mammograms, with the aim of detecting breast cancer at an early stage.

Several studies have shown promising results, with machine learning algorithms achieving high accuracy rates in detecting early-stage breast cancer.

For example, a recent study published in the Journal of the National Cancer Institute found that a machine learning algorithm was able to identify breast cancer in mammograms with an accuracy of 90%, which is comparable to that of trained radiologists.

Moreover, the development of deep learning algorithms has shown even greater potential in detecting early-stage breast cancer. Deep learning algorithms can automatically learn and identify patterns in data and have been shown to achieve high levels of accuracy in detecting breast cancer in mammograms.

While the use of machine learning in breast cancer detection is still in its early stages, there is a great potential for these techniques to improve the accuracy and efficiency of breast cancer screening, ultimately leading to better patient outcomes and reduced healthcare costs.

The research subject of the present study is the development of machine learning algorithms using physical laws for approximating the resulting dataset. The following objectives need to be achieved for this purpose:

- Determining the most suitable neural network architectures for studying the behavior of non-linear partial differential equations.
- Developing deep learning methods for solving direct and inverse problems.
- Evaluating the effectiveness of training and the solution system of the developed neural network.

The scientific novelty of the research lies in the development of the existing methodology for statistical analysis and forecasting of time series. The main theoretical and practical results that constitute the subject of protection and scientific contribution are as follows:

1) Compared to other, more traditional ways of solving machine learning problems, physical information algorithms significantly reduce time and costs by training only a certain part of the final model.

2) Criteria for selecting a scientifically justified forecasting method have been identified, based on the accuracy of results, cost, expert assessment, and other context-dependent characteristics.

Expected results: A machine learning model for more accurate and efficient early detection of breast cancer.

The practical significance of the study lies in the fact that its provisions and conclusions can be used in further, more in-depth research of statistical methodology. The theoretical and methodological provisions presented in the thesis significantly improve the quality of time series analysis and forecasting, refine the peculiarities of method implementation, and provide a deep understanding of the essence of the underlying processes. The developed physically informed neural network model has practical relevance in terms of its predictive functions and the ability to identify existing causal relationships in this field.

To adapt this text for the topic of breast cancer detection using machine learning techniques, it could be revised as follows:

This study aims is about developing machine learning algorithms to improve the accuracy of early detection of breast cancer. The following objectives need to be achieved during diploma work:

– Identifying the most effective neural network architectures for analyzing large datasets of mammograms and detecting breast cancer at an early stage.
– Developing deep learning methods for analyzing medical images and achieving high levels of accuracy in breast cancer detection.
– Evaluating the effectiveness of the developed neural network in diagnosing breast cancer.

The practical significance of the study lies in the potential to improve the accuracy and efficiency of breast cancer screening programs, ultimately leading to better patient outcomes and reduced healthcare costs.

Results. In the course of our work, we used a dataset from Kaggle, it contains about 30 columns that describe the disease. For example: the size of the tumor, the type of tumor (malignant or benign), the texture of the tumor, and many other characteristics. According to these characteristics, more than 500 records have been recorded.

In this study, five-level master data exploration approaches using five different machine learning prediction models, including SVM, logistic regression, KNN, random forest, and decision tree classifier, were proposed to identify and classify breast cancer tumors.

Meanwhile, the random forest achieved an accuracy of 92.55%, demonstrating that machine learning algorithms may identify higher precision. Our findings demonstrate the accuracy of our prediction models for detecting breast cancer while requiring only a short model training period of time. These complex models, methodologies, and findings will assist physicians and data analysts in using a more intelligent classifier to identify breast cancer symptoms.

Since breast cancer-related picture data is accessible, we applied deep learning models for breast cancer diagnosis - ANN, which resulted in an excellent outcome of 100 epochs with 98% accuracy.

During the analysis, we focused on how the model was trained and how the data sets matched the final outcome. We tried several options and came to the conclusion that this model can predict the type of breast cancer tumor from the data with 98% accuracy, which is an excellent result. This model can serve as an assistant for doctors, analysts, it can also be modified for datasets related to medicine,

The results of the study indicate that deep learning models, specifically physical-informed neural networks, show promise in providing accurate predictions for a variety of applications, such as forecasting physical processes based on data, model prediction control, multi-physics and multi-scale modeling, and simulation. However, questions remain regarding the optimal architecture and data requirements of these models, as well as the potential limitations of using them in certain scenarios. The study also highlights the importance of integrating classical numerical methods with deep learning approaches, as both can bring valuable insights to the field of computational physics.

Overall, this work contributes to the ongoing synergy between machine learning and computational physics and has the potential to lead to innovative and efficient developments in both areas.