InterConf
Scientific Publishing Center

ISSN 2709-4685

THE ISSUE CONTAINS:

Proceedings of the 8th International Scientific and Practical Conference

SCIENTIFIC TRENDS AND TRENDS IN THE CONTEXT OF GLOBALIZATION

Umeå, Kingdom of Sweden
19-20.11.2024

Scientific Collection
InterConf+

№ 51 (225)
November, 2024

OPEN ACCESS

ISSN 2709-4685

Scientific Collection «InterConf+»

№ 51(225)

November, 2024

THE ISSUE CONTAINS:

Proceedings of the 8th International
Scientific and Practical Conference

SCIENTIFIC TRENDS AND TRENDS IN THE CONTEXT OF GLOBALIZATION

UMEÅ, KINGDOM OF SWEDEN

November 19-20, 2024

UMEÅ
2024

ICV (Index Copernicus Value) = 88.21 / 100

UDC 001.1

S 40 *Scientific Collection «InterConf+»*, 51(225): with the Proceedings of the 8th International Scientific and Practical Conference «Scientific Trends and Trends in the Context of Globalization» (November 19-20, 2024; Umeå, Kingdom of Sweden) / comp. by LLC SPC «InterConf». Umeå: Mondial, 2024. 548 p.

ISSN 2709-4685
DOI 10.51582/interconf.19-20.11.2024

Please, cite as shown below:

1. Surname, N. & Surname, N. (2024). Title of an article. *Scientific Collection «InterConf+»*, 51(225), 21-27. https://doi.org/10.1080/interconf...

This issue of Scientific Collection «InterConf+» contains the materials of the International Scientific and Practical Conference. The conference provides an interdisciplinary forum for researchers, practitioners and scholars to present and discuss the most recent innovations and developments in modern science. The aim of conference is to enable academics, researchers, practitioners and college students to publish their research findings, ideas, developments, and innovations.

Scientific Collection «InterConf+» and its content are indexed in:

Index Copernicus; Google Scholar; WorldCat; OUCI (*Open Ukrainian Citation Index*); CrossRef; Semantic Scholar; Mendeley; Scilit; OpenAIRE (*pan-European research information system*), etc.

TABLE OF CONTENTS

LAW AND INTERNATIONAL LAW

ARTS, CULTURAL STUDIES AND ETHNOGRAPHY

HISTORY AND ARCHEOLOGY, ARCHIVAL STUDIES

MEDICINE AND PHARMACY

NATURE MANAGEMENT, RESOURCE SAVING AND ECOLOGY

ENERGETICS

PHYSICS AND MATHS

CHEMISTRY AND MATERIALS SCIENCE

AGROTECHNOLOGIES AND AGRICULTURAL INDUSTRY

GENERAL ENGINEERING AND MECHANICS

INFORMATION AND WEB TECHNOLOGIES

PHYSICAL EDUCATION AND SPORTS

Proceedings of the 8th International Scientific and Practical Conference «Scientific Trends and Trends in the Context of Globalization»

(November 19-20, 2024). Umeå, Kingdom of Sweden

No 225

INFORMATION AND WEB TECHNOLOGIES

DOI 10.51582/interconf.19-20.11.2024.050

Development and estimation of the models for early prediction of pancreatic cancer using deep learning

Shmatko Oleksandr[1], Gamayun Igor[2], Dorogiy Mikola[3]

[1] PhD, Associate Professor;
Technical University «Metinvest Polytechnic» LLC; Ukraine

[2] Doctor of Technical Science, Professor;
National Technical University «Kharkiv Polytechnic Institute»; Ukraine

[3] Masters Student;
National Technical University «Kharkiv Polytechnic Institute»; Ukraine

Abstract.
The development of efficient diagnostic tools for the early detection of pancreatic cancer is critical due to the disease's high mortality rate and often late diagnosis. This research focuses on leveraging deep learning models, particularly convolutional neural networks (CNN) and YOLO-based architectures, to enhance the accuracy of cancer detection using medical imaging. The article reviews existing techniques and proposes a novel approach to image classification for identifying cancerous growths in the pancreas. The results demonstrate improved accuracy in diagnosis, highlighting the potential for early intervention.

Keywords:
Pancreatic cancer
deep learning
convolutional neural networks
medical imaging
YOLO

Proceedings of the 8th International Scientific and Practical Conference «Scientific Trends and Trends in the Context of Globalization»

(November 19-20, 2024). Umeå, Kingdom of Sweden

No 225

INFORMATION AND WEB TECHNOLOGIES

Introduction. The importance of disease classification and prediction has become increasingly evident in recent years. To accurately identify the causes and symptoms of a disease, it is essential to have a deep understanding of the relevant features in the available datasets. Artificial intelligence (AI) has shown promising results in classifying diseases and assisting in decision-making processes. Within AI, machine learning (ML) has accelerated numerous medical research projects, with deep learning (DL), a subset of ML, focusing on multi-layered neural networks to analyze critical features necessary for disease detection [1, 2, 3]. Since 2014, numerous studies have discussed the application of algorithms developed to advance the medical field, providing more accurate results for patients. AI, through the analysis of data, has stimulated the development of advanced technologies in areas such as natural language processing, speech recognition, and computer vision, enabling the creation of reliable systems like autonomous vehicles and automated translation systems. Despite these achievements, the application of AI in medicine still faces challenges, particularly related to the risks and difficulties associated with ensuring accurate predictions based on collected medical data and managing complex healthcare systems.

AI processes datasets by extracting necessary features or identifying patterns from vast amounts of data, which can present difficulties in isolating the most meaningful features and excluding redundant ones. These challenges make it difficult to maintain accuracy in prediction metrics. Consequently, selecting a smaller subset of relevant features from a large set can improve model efficiency. By eliminating redundant and irrelevant features, the dimensionality of the dataset is reduced, which accelerates model training, similar to boosting methods [4]. Common techniques such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are used to extract significant features. Feature selection serves two main purposes: first, to improve data representation, and second, to reduce the number of features, addressing the issue of high dimensionality. Thus, feature selection plays a crucial role in achieving these objectives. Further studies have expanded feature selection strategies by

INFORMATION AND WEB TECHNOLOGIES

incorporating multi-objective optimization methods. This review will focus on strategies for effective feature selection.

Cancer has been detected using various image segmentation, feature selection, and regression methods, including Root Mean Square Error (RMSE) for evaluating parameters such as pattern recognition, object detection, and image classification [5]. Brain tumors, for example, have been detected using six classifiers and Transfer Learning (TL) methods to segment brain magnetic resonance imaging (MRI) scans [6]. Furthermore, TL has been applied in identifying lung cancer and brain diseases, as discussed in [7], where MRI and computed tomography (CT) images were analyzed using Support Vector Machines (SVM). Although image analysis processes are well-studied in existing research, methods that employ ML and DL are continuously evolving, which presents challenges for researchers seeking to identify the most effective image analysis techniques and feature extraction methods. This study makes several contributions:

- Classification of diseases based on a review of primary research studies,
- Recognition of various image modalities discussed in existing literature,
- Description of tools and reliable ML and DL methods for disease prediction,
- A review of datasets to raise awareness about available data sources,
- Experimental results comparing various ML and DL methods using MRI datasets,
- Selection of relevant features and classifiers to achieve better accuracy,
- Insights into classification techniques and suggestions for future research directions.

The significance of this review lies in providing clinicians with the tools to use ML or DL methods for accurate and reliable disease detection, classification, and diagnosis. It will also help clinicians and researchers avoid misinterpretation of datasets, derive efficient algorithms for disease diagnosis, and offer insight into modern medical imaging techniques that employ ML and DL.

INFORMATION AND WEB TECHNOLOGIES

Literature Review. Machine Learning (ML) is a subfield of artificial intelligence (AI) where machines learn from data by identifying patterns and automating decision-making processes with minimal human intervention [9, 10, 11]. One of the key characteristics of ML models is their ability to autonomously adapt, learn from previous computations, and provide reliable results when exposed to new datasets. There are two primary aspects to ML's contribution in the medical field:

-ML methods assist physicians in interpreting medical images through computer-aided design (CAD) systems within a short timeframe.

-Algorithms are employed for complex tasks such as segmentation in CT scans [12], breast cancer detection via mammography, and brain tumor segmentation using MRI.

Traditional ML models worked with structured datasets where the methods were predefined for each step, and the applied method would fail if any step was omitted. The evaluation of data quality used by ML and Deep Learning (DL) algorithms is critically important [13-16]. Newer algorithms have adapted to handle missing data based on the robustness requirements of the model. DL models, in particular, enable machines to achieve higher accuracy by refining methods for analyzing medical images. For example, in [17], heart disease was diagnosed using labeled chest X-rays, where a cardiologist reviewed and re-labeled the data, excluding irrelevant images. Data augmentation and Transfer Learning (TL) were used to extract precise features, resulting in 82% accuracy, 74% specificity, and 95% sensitivity for heart failure. Similarly, in [18], automatic feature selection was developed using histopathological images marked as either positive or negative for cancer, with minimal manual intervention. Two networks, Deep Neural Network (DNN) 2-F and DNN1-F, were employed with Principal Component Analysis (PCA) to reduce features in the DNN, while a single-layer K-means centroid network was used for unsupervised feature learning. Later, the results of unsupervised (93.56%) and supervised (94.52%) learning were compared. DL models automate the feature extraction process to efficiently handle data [19, 20]. Figure 1.2 illustrates the process used by DL algorithms for

INFORMATION AND WEB TECHNOLOGIES

predicting and diagnosing various diseases.

Both ML and DL methods have been utilized for processing medical images to improve prediction and accuracy, as shown in Figures 1.1 and 1.2. Medical images in various formats serve as input data, which are then subjected to algorithms. The input image is segmented based on various factors, and key features are extracted from these segments using feature extraction methods. After extracting the necessary features, they are refined to identify the actual features used for disease detection [21]. Moreover, in [22], ML approaches were employed to denoise medical images, improving prediction and accuracy. After feature extraction and noise removal, image classification based on the disease was achieved using classifiers such as Support Vector Machines (SVM), decision trees (DT), etc.

ML is a process in which computers learn from data and use algorithms to perform tasks without being explicitly programmed. It utilizes pattern recognition to make predictions on new datasets. In contrast, DL is modeled after the human brain, incorporating a complex structure of algorithms that allows machines to process images, text, and documents. DL employs multi-layered algorithms such as Convolutional Neural Networks (CNN), Artificial Neural Networks (ANN), and others to analyze data using logical processes. Compared to ML models, DL is capable of processing larger volumes of data.

This section reviews articles related to the symptoms, detection, classification, prediction, and diagnosis of breast cancer using ML and DL methods. In [23], the significance of using the BI-RADS (Breast Imaging Reporting and Data System) for developing a CAD system for acquiring breast ultrasound images was discussed. Furthermore, a 10-fold cross-validation technique was used for benign and malignant tumor classification, achieving 77% accuracy with the SVM classifier. However, some methods involving multiple algorithms that handle large volumes of data require a deeper understanding and analysis [24]. CNN was used to train the system on clinical data and to understand its complex structure. Additionally, a study was proposed to explore radiomics and CADx expansion for extracting tumor features

using a CAD system. Breast cancer was classified based on parameters such as Area Under the Curve (AUC), sensitivity, and specificity [100]. The CAD system was developed using CNN, requiring numerous features, including multi-angle features, to maximize the detail in image data for accurate detection and classification.

DL has been used for analyzing medical images, and the limitations and successes of DL methods for medical imaging were discussed in [25]. Various ML methods used for image processing, along with DL techniques and algorithm architecture, were examined. Histological images, thermographic images, mammography, ultrasound, and MRI were analyzed using the CAD system. The system incorporated ML methods such as SVM, ANN, DT, Naive Bayes, and K-Nearest Neighbor (KNN), among others.

Main part.

To improve the accuracy of our results, we utilized the standard YOLOv4 design, where DarkNet53 serves as the backbone of the network and a three-layer spatial pyramid is implemented as the neck. The Binary Cross Entropy (BCE) loss function was applied as the target loss function in the detector head, and the original YOLOv4 implementation was further optimized by adding branching and loss functions.

The YOLOv4 model was employed for both cancer detection as an object and image classification. Given that classification accuracy was more crucial for early cancer detection than the detection region, we assigned greater weight to classification loss. We initiated the network configuration by generating parameters randomly. This approach was used to ensure that the activation function data for each layer at the early stage of training fell within a reasonable range, which was essential for guaranteeing the network's quick convergence.

Since the dataset we used was significantly smaller than the dataset typically used in the YOLOv4 network, there was a risk of overfitting if we strictly adhered to the provided boundaries. To mitigate this, we first conducted preliminary training on the DarkNet53 backbone for image recognition tasks in ImageNet, as well as object classification tasks within the dataset. Following this, we added the three-layer pyramid

INFORMATION AND WEB TECHNOLOGIES

neck for detection and fine-tuned it using a dataset specifically designed for early-stage cancer detection. As illustrated in Figure 1, we normalized the images from ImageNet and the source data by adjusting the range and mean values to best align the learning rate with the early tumor dataset. This step allowed us to optimize model parameters for the early-stage cancer dataset.

For each iteration of the fine-tuning process, we used 64 images sized 224x224 pixels. Due to the limited memory available on the GPU, the batch was divided into 32 parts. A total of 100 epochs were conducted, with the first two epochs serving as a warm-up phase and using a cosine learning rate of 0.01 per epoch. After the warm-up phase, the learning rate was decreased to 0.001 for each subsequent epoch.

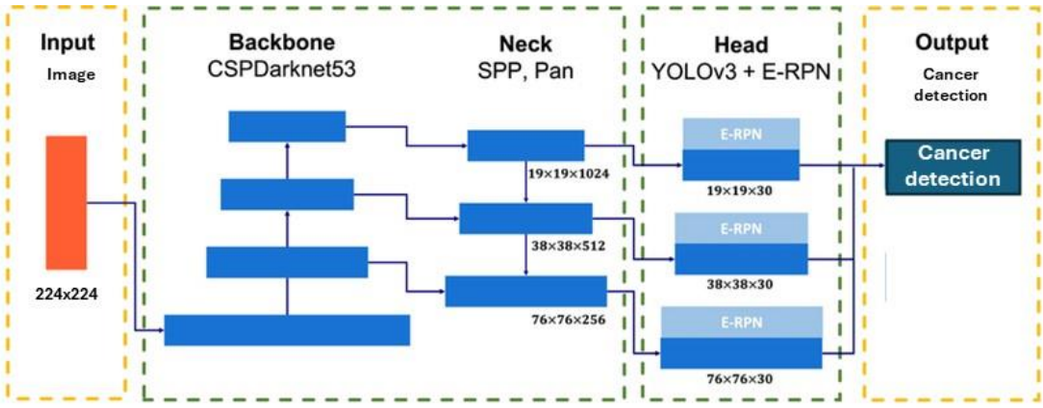

Figure 1

Example of the YOLO architecture used in the proposed model

All experiments were conducted on a 64-bit computer equipped with an Intel(R) Core(TM) i3-10110U CPU @ 2.10 GHz 2.59 GHz processor and 8 GB of RAM. For training and validating the model, the code was implemented in Python using the Google Colab platform.

The image dataset was collected between June 2017 and June 2018. The dataset comprises 3,494 Computed tomography (CT) scans from 222 patients with pathologically confirmed pancreatic cancer, while a control dataset of 3,751 CT scans was collected from 190 healthy individuals. Based on these images, a Convolutional Neural Network (CNN) model was

INFORMATION AND WEB TECHNOLOGIES

developed. Examples of the images are shown in Figure 2.

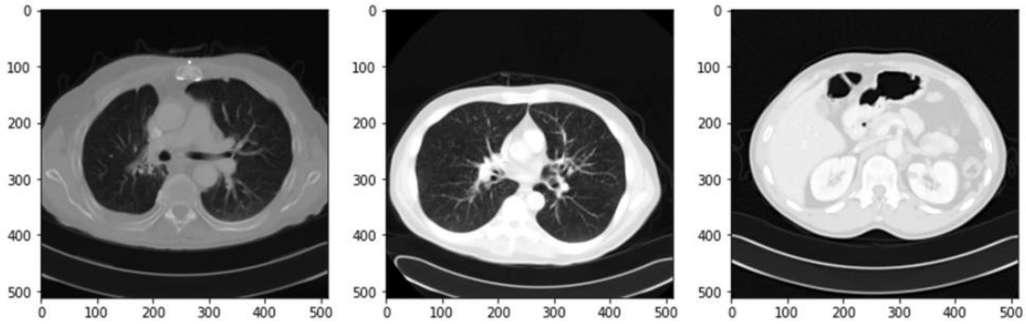

Figure 2
Image Samples

The images were divided into three datasets based on different phases, and the proposed method was tested using tenfold cross-validation for binary classification (i.e., cancer or no cancer). Figure 3 illustrates the accuracy of pancreatic cancer detection within the image dataset. The proposed model demonstrated nearly 100% accuracy in classifying pancreatic cancer.

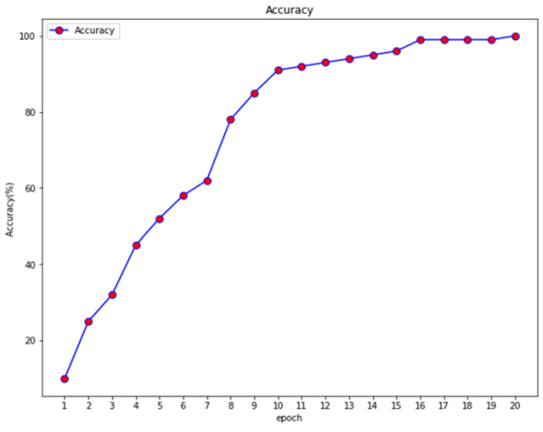

Figure 3
Accuracy of Cancer Detection in the Image Dataset

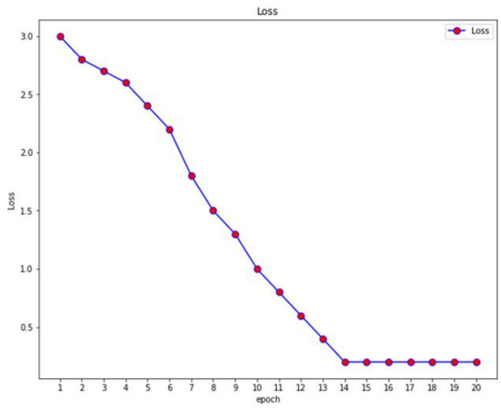

Figure 4
Loss Function

Given that accuracy evaluates the classifier's performance across all classes, not just a single class Ci,

INFORMATION AND WEB TECHNOLOGIES

it was the primary metric used to assess the effectiveness of our method. Additionally, the loss function was evaluated, and the values are presented in Figure 4.

Figure 5 shows the confusion matrix for the predictions related to non-cancerous and cancerous outcomes in the image dataset. The confusion matrix confirms that the model achieves 100% accuracy on the CT image dataset.

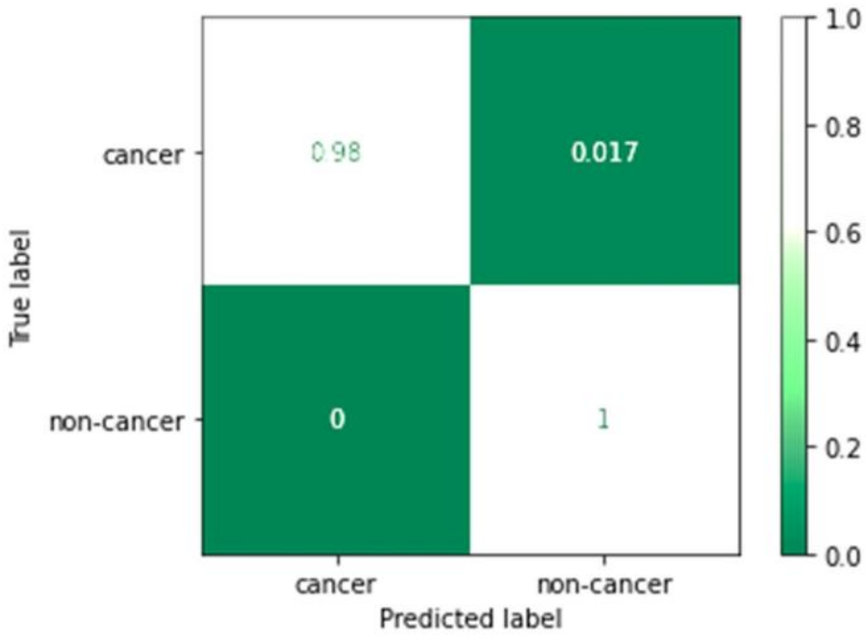

Figure 6

Confusion Matrix for the Image Dataset

Table 1 provides the precision, recall, and F1 score obtained from our proposed model. The results indicate that the proposed model achieves 100% classification accuracy, demonstrating its robustness in detecting pancreatic cancer.

Table 1

Values of model evaluation criteria's

Criteria	Precision	Recall	f1-score
1	100%	100%	100%
2	98%	100%	99%
3	100%	99%	98%

INFORMATION AND WEB TECHNOLOGIES

Table continuation 1

Accuracy	100%		
Macro avg	100%	100%	100%
Weighted avg	100%	100%	100%

This approach enabled us to balance the complexity of the model with the size of the dataset, thereby ensuring the accurate detection of early-stage cancer with minimal risk of overfitting.

Conclusions. This research highlights the effectiveness of CNN models in the early detection of pancreatic cancer. By employing a YOLO-based approach, the system was able to identify cancerous tissues with high precision, offering a potential solution for earlier and more reliable diagnosis. Future work will focus on enhancing the model's capability through the integration of additional medical imaging data and further refining the segmentation techniques. This approach not only improves diagnostic accuracy but also paves the way for integrating deep learning models into real-time medical diagnostic tools.

References:

[1] Abdulbaqi AS, Younis MT, Younus YT, Obaid AJ (2022) A hybrid technique for EEG signals evaluation and classification as a step towards neurological and cerebral disorders diagnosis. Int J Nonlinear Anal Appl 13(1):773-781

[2] Aceto G, Persico V, Pescapé A (2018) The role of information and communication technologies in healthcare: taxonomies, perspectives, and challenges. J Netw Comput Appl 107:125-154. https://doi.org/10.1016/j.jnca.2018.02.008

[3] Acharya UR, Fujita H, Oh SL, Hagiwara Y, Tan JH, Adam M (2017) Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. Inf Sci 415:190-198. https://doi.org/10.1016/j.ins.2017.06.027

[4] Aggarwal 1LP (2019) Data augmentation in dermatology image recognition using machine learning. Skin Res Technol 25(6):815-820. https://doi.org/10.1111/srt.12726

[5] Al-Najdawi N, Biltawi M, Tedmori S (2015) Mammogram image visual enhancement, mass segmentation and classification. Appl Soft Comput 35:175-185. https://doi.org/10.1016/j.asoc.2015.06.029

[6] Altan G, Kutlu Y, Allahverdi N (2019) Deep learning on computerized analysis of chronic obstructive pulmonary disease. IEEE J Biomed Health Inf 24(5):1344-1350. https://doi.org/10.1109/JBHI.2019.2931395

INFORMATION AND WEB TECHNOLOGIES

[7] Anbeek P, Vincken KL, Van Bochove GS, Van Osch MJ, van der Grond J (2005) Probabilistic segmentation of brain tissue in MR imaging. NeuroImage 27(4):795-804. https://doi.org/10.1109/TMI.2014.2366792

[8] Arya R, Kumar A, Bhushan M (2021) Affect recognition using brain signals: a survey. In: Computational methods and data engineering. Springer, Singapore, pp 529-552. https://doi.org/10.1007/978-981-15-7907-3_40

[9] Arya R, Kumar A, Bhushan M, Samant P (2022) Big five personality traits Prediction using brain signals. Int J Fuzzy Syst Appl (IJFSA) 11(2):1-10. https://doi.org/10.4018/IJFSA.296596

[10] Noisy Data in Data Mining | Soft Computing and Intelligent Information Systems (ugr.es) Available at: https://techblog.cdiscount.com/a-brief-overview-of-automatic-machine-learning-solutions-automl/

[11] Beam AL, Kohane IS (2018) Big data and machine learning in health care. JAMA 319(13):1317-1318. https://doi.org/10.1001/jama.2017.18391

[12] Bhatt C, Kumar I, Vijayakumar V, Singh KU, Kumar A (2021) The state of the art of deep learning models in medical science and their challenges. Multimedia Syst 27(4):599-613. https://doi.org/10.1007/s00530-020-00694-1

[13] Bhattacharya P, Tanwar S, Bodkhe U, Tyagi S, Kumar N (2019) Bindaas: Blockchain-based deep-learning as-a-service in healthcare 4.0 applications. IEEE Trans Netw Sci Eng 8(2):1242-1255

[14] Bhattacharyya A, Pachori RB, Upadhyay A, Acharya UR (2017) Tunable-Q wavelet transform based multiscale entropy measure for automated classification of epileptic EEG signals. Appl Sci 7(4):385

[15] Bhushan M, Goel S (2016) Improving software product line using an ontological approach. Sādhanā 41(12):1381-1391

[16] Bhushan M, Goel S, Kumar A, Negi A (2017) Managing software product line using an ontological rulebased framework. In: 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions)(ICTUS). IEEE, pp 376-382

[17] Bhushan M, Goel S, Kaur K (2018) Analyzing inconsistencies in software product lines using an ontological rule-based approach. J Syst Softw 137:605-617

[18] Bhushan M, Goel S, Kumar A (2018) Improving quality of software product line by analysing inconsistencies in feature models using an ontological rule-based approach. Expert Syst 35(3):e12256

[19] Bhushan M, Negi A, Samant P, Goel S, Kumar A (2020) A classification and systematic review of product line feature model defects. Software Qual J 28(4):1507-1550

[20] Bhushan M, Kumar A, Samant P, Bansal S, Tiwari S, Negi A (2021) Identifying quality attributes of FODA and DSSA methods in domain analysis using a case study. In: 2021 10th international conference on System Modeling & Advancement in Research Trends (SMART). IEEE, pp 562-567

[21] Bhushan M, Duarte JÁG, Samant P, Kumar A, Negi A (2021) Classifying

INFORMATION AND WEB TECHNOLOGIES

and resolving software product line redundancies using an ontological first-order logic rule based method. Expert Syst Appl 168:114167

[22] Caballé-Cervigón N, Castillo-Sequera JL, Gómez-Pulido JA, Gómez-Pulido JM, Polo-Luque ML (2020) Machine learning applied to diagnosis of human diseases: a systematic review. Appl Sci 10(15):5135. https://doi.org/10.3390/app10155135

[23] Cabitza F, Rasoini R, Gensini GF (2017) Unintended consequences of machine learning in medicine. JAMA 318(6):517-518. https://doi.org/10.1001/jama.2017.7797

[24] Caliskan A, Badem H, Basturk A, YUKSEL M (2017) Diagnosis of the parkinson disease by using deep neural network classifiers. IU-J Electr Electron Eng 17(2):3311-3318

[25] Chaganti SY, Nanda I, Pandi KR, Prudhvith TG, Kumar N (2020) Image classification using SVM and CNN. In: 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA). IEEE, pp 1-5

SCIENTIFIC EDITION

SCIENTIFIC COLLECTION «INTERCONF+»

№ 51(225) | November, 2024

The issue contains:

Proceedings of the 8th International
Scientific and Practical Conference

SCIENTIFIC TRENDS ANDTRENDS IN THE CONTEXT OF GLOBALIZATION

Umeå, Kingdom of Sweden
19-20.11.2024

All materials are reviewed.
The editorial office did not always agree with the position of authors.

Journal's frequency: monthly

Sighed for online publication: November 20, 2024.

Printed: December 19, 2024. Circulation: 200 copies. Format 60×84/8.
Batang & Courier New typefaces. Offset paper 100gsm. Digital color printing.

Contacts of the editorial office:

LLC Scientific Publishing Center «InterConf»
info@interconf.center
https://www.interconf.center

Certificate on the entry of publishing business subject in the State Register of Publishers, Manufacturers and Distributors of Publishing Products of Ukraine: ДК № 7882 of 10.07.2023.