International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A System for Decision-Making Assistance Using Human–Computer Interactions for Cancer Detection

[1]Prerna Mahajan, [2]Sunil Sharma, [3]Dr. Simarjeet Makkar, [4]Pooja Sharma, [5]Preeti Naval

Submitted: 10/01/2024 **Revised**: 16/02/2024 **Accepted**: 24/02/2024

Abstract: The increasing growth of telemedicine and recent developments in diagnostic artificial intelligence (AI) make it essential to think through the benefits and drawbacks of incorporating AI-based assistance into novel healthcare approaches. In this study, we expand on previous advances in the precision of image-based AI for skin cancer (SC) detection to consider the consequences of various presentations of AI-based assistance across varying degrees of clinical knowledge and various clinical processes. We search that less-experienced clinicians benefit the most from AI-based assistance of clinical decision-making (CDM) and that AI-based support enhances diagnostic precision above that of either AI. We also discover that in the context of mobile technologies, AI-based multiclass probabilities outperform content-based image retrieval (CBIR) presentations of AI. Together, our method and outcome present a basis for further study across the whole range of image-based diagnostics, with the ultimate goal of enhancing human-computer cooperation in clinical settings.

Keywords: *Artificial intelligence, content-based image retrieval, cancer detection, decision making.*

1. Introduction

In the process of developing health information technology (IT), such as "electronic health records (EHRs) and computerized clinical decision support (CDS)", the part of "human factors and human-computer interaction (HCI)" technologies is underutilized. Incorporating these strategies into these products may enhance their usefulness in different ways, such as a reduction in the amount of labor required and an increase in the level of pleasure experienced by users. There are field-based and laboratory-based methods that are used in the discipline of human factors and ergonomics, which is a subfield within the study of human-computer interaction (HCI). Every strategy comes with its own unique set of benefits and drawbacks [1]. Methods used in the field, including fast ethnography and conversations, can both retain the surroundings of the workplace in which health IT is applied and gather the diverse nature of that setting. On the other hand, techniques that are conducted in a laboratory provide the opportunity to manipulate experimental conditions and evaluate dependent variables in an environment that is tightly controlled [2]. It is anticipated that cancer will soon overtake cardiovascular disease as the top cause of mortality globally, accounting for one death in every six that occur. Therefore, increasing cancer prediction becomes crucial for improving survival rates, which may be accomplished by giving chances for early-diagnosed individuals to get therapy that is suitable for their condition. There have already been uses of microarray technology in the healthcare industry to enhance the prediction of cancer. Analyses of DNA microarray data and analyses of biomarkers may both be performed concurrently by medical professionals and biologists to diagnose and/or forecast different forms of cancer [3].

Image-based AI has the potential to increase the precision of visual diagnostics. The majority of research that has been done up to this point has been based on direct precision of diagnosis studies comparing AI-based technologies to human practitioners. In a similar vein, recent research in the field of dermatology has shown that AI for some types of skin lesions is either comparable to or even more accurate than human specialists when it comes to image-based diagnosis when conducted under controlled settings [4].

It has not yet been determined what function human–computer cooperation plays in the provision of medical care, which circumstances are most suitable for its use, or how it affects the quality of medical treatment received. To examine the consequences of different presents of AI-based aid across several healthcare procedures and diverse degrees of professional expertise, we looked at the use case of an SC diagnosis. The images of AI that we choose to adopt originated from the existing body of scholarly work and vary significantly in crucial features such as simplicity, granularity, and concreteness [5]. Since we were tasked with solving a classification issue involving many classes, one apparent option was to give AI-based probabilities for multiple classes. The second approach was sparked by

[1]Jain (Deemed to be University), Bangalore, Karnataka, India, Email Id- prerna.m@jainuniversity.ac.in, Orcid Id- 0000-0002-2408-024X
[2]Vivekananda Global University, Jaipur sunil.sharma@vgu.ac.in, Orcid Id- 0009-0004-5850-951X
[3]Department of ISME, ATLAS SkillTech University, Mumbai, Maharashtra, India, Email Id- simarjeet.makkar@atlasuniversity.edu.in, Orcid Id- 0009-0006-6364-8201
[4]Chitkara University, Rajpura, Punjab, India pooja.sharma.orp@chitkara.edu.in, https://orcid.org/0009-0005-5217-1711
[5]Maharishi University of Information Technology, Lucknow, India, Email Id- er.preetinaval09@gmail.com, Orcid Id- 0000-0003-2988-7082

problems addressed by existing AI-based support for SC screening. We showed the AI-predicted probability of malignancy and then dichotomized the disease classes into benign and malignant classes. This technique was successful in reducing the number of false positive diagnoses. Using the same Deep learning (DL) algorithm, we developed a sort of AI-based CBIR that assists medical professionals in the interpretation of images by searching databases to find comparable images that have been labeled with known illnesses. This was the third method, which was fundamentally distinct from the previous two [6].

This research takes a multifaceted look at human-computer teamwork in a range of contexts. For brevity's sake, we focused on SC identification; nevertheless, our work might provide the groundwork for comparable investigations into other areas of image-based diagnostic medicine. Our research contradicts the prevalent narrative by arguing that human-computer cooperation, rather than rivalry, should be prioritized. In terms of regulation, the effectiveness of AI-based solutions should be evaluated not in laboratory settings but in the hands of end users. Only then can we hope to use AI-based decision assistance logically, develop it, and speed up its progress.

The document is structured as shown in the following example. In Section 2, a thorough analysis of the cancer detection is provided. The entire description of the proposed method may be found in Section 3. In Section 4, along with a discussion, the implementation's results are shown. The key contributions of the article are examined in Section 5, along with the possibility of further development of the proposed program.

2. Related Work

In [7], the information requirements of medical specialists for the implementation of an AI Assistant were outlined. There has been a lot of research done on the issue of explaining model predictions, but this study implies that it may also be helpful to give transparency into the higher-level design aims of the model itself, as well as its global behavior and trends. It is possible for machine learning (ML)-based AI systems [8] trained on inaccurate or inadequate data to exhibit biassed "thinking," which might amplify prejudice and inequity, disseminate false information, or even result in physical damage. These worries prevented several ambitious AI initiatives from ever getting off the ground. Human lives will become more impacted by choices affected by AI thinking and worldviews as more and more corporate and government services based on AI/ML algorithms become available. To guarantee that ethical AI solutions put the user and human values first, HAI research [9] is crucial. Experts from the social sciences and moral philosophy must work together to tackle the obstacles that make HAI such an important project. Due to its distinct nature, [10] interventional cardiology is a prime candidate for the introduction of AI-based technologies that aim to enhance real-time CDM, simplify the catheterization laboratory's workflow, and standardize catheter-based procedures through cutting-edge robotics. Articles in [11] are organized around six topics such as “Interfaces, Visualization, Electronic Health Records, Devices, Usability, and CDS Systems” and evaluate and debate a variety of research disciplines directly related to HCI and CDM. These publications, however, often have overlapping themes, demonstrating how HCI links together otherwise unrelated fields of study. The papers under examination were divided into four categories so that HCI and design issues could be more closely examined. Our [12] findings provide the encouraging message that most managers are open to working in tandem with robots, so long as humans maintain the impression that choices are based mostly on their input and judgment. Still, one must proceed with care since many vigorous objects to the use of intelligent computers with administrative authority.

3. Proposed Methodology

HCI describes the two-way flow of data between computers and humans, whereby the latter receives information from the former through input and output devices. The most significant material, as indicated by communicative behavior, is simulated in a multimodal simulation, which is a tangible 3D virtual realization of the situational environment and coexisting agent. DL is a subfield of ML that makes use of artificial neural networks to model the challenging multilayered structure and behavior of the human brain. HCI-DL technology, enabled by robust computers, large datasets, and sophisticated algorithms, has proven effective in a variety of IT applications across a broad range of sectors and fields. Image interpretation, radiology information distribution and application, CDM, and outcome monitoring are just a few of the areas where this technology is anticipated to have a profound effect. Specific frequencies and cancerous "alert" noises were used to indicate the cluster centroids, which were then transferred into the volume, timbre, and length of a sonification. As a consequence, audio signals were created for each of the data centroids. These audio signals allowed for the acoustic separation of the benign from malignant lesions and supplied data on the image via a raw wave file, as seen in figure 1.

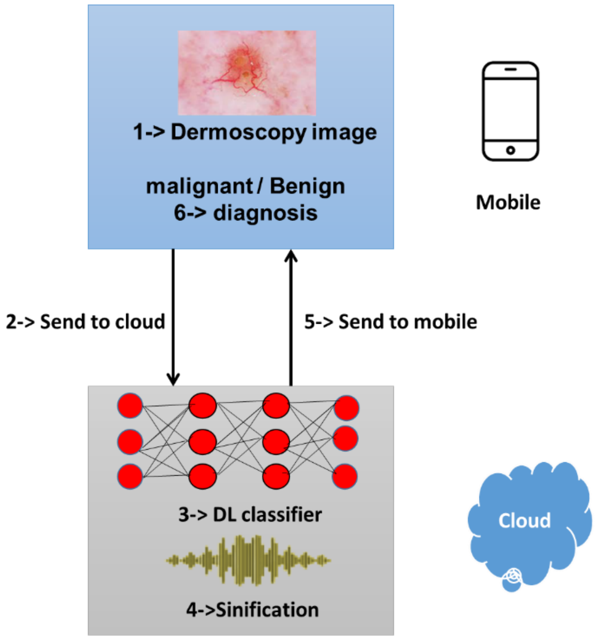

Fig. 1. Dermoscopy image

3.1 Dataset

The HAM10000 dataset was used to fine-tune a DL for classification purposes. We used the Pytorch framework to train on NVIDIA GPUs; our architecture was a ResNet34; we seeded the weights using data from ImageNet [13]. As a loss function, we used cross-entropy, with weights that varied according to how often each class appeared in the dataset. The initial learning rate was 0.0001, and it was set to drop to 1111 if the validation loss didn't improve after three iterations. For this reason, we only ran up to 102 training epochs before cutting off further optimization. The 34 images in each batch were randomly sized to 224226 pixels without normalization to a mean pixel, spin by 92^{o}, and then turned with a little amount of colour, contrast, saturation, and hue jitter.

Images for training and 5-fold cross-validation came from the publicly accessible HAM10000 dataset, which matches the ISIC 2018 challenge's training set. When it came to the hold-out validation set, we narrowed it down to a single network that performed well. The photos were reduced to 224x224 pixels and cropped to 80% for inference; for the sake of the tests, the images were then flipped horizontally and rotated by either 0 or 90 degrees. The colour was also normalized for the telemedicine dataset using Shades of Grey and a Minkowski norm of 6. After using a softmax function to include all class possibilities from 0 to 100%, the resulting multiclass probabilities were given to the raters. By running the same DL on both the target and the HAM10000 datasets, we were able to extract the feature vector of the target image and apply cosine similarity to locate other similar photos. In the CBIR decision support powered by AI, we saved the four most similar photos for each class and displayed them[16].

3.2 Platform for Interaction and Raters

Online communication tool: The Medical University of Vienna's DermaChallenge website served as the interface for evaluating the efficacy of both human and AI raters. The platform is made up of a back end and a front end and is based on a stack of commonly used web technologies. The back end supports JavaScript Object Notation web tokens for user authentication and a presentational state transfer interface for loading and persisting data. All data transmissions are encrypted using the TLS protocol via an SSL connection. The front end has been designed with mobile devices in mind, although it is accessible from any system with a web browser that supports JavaScript. Five people tried out the system before it was released to the public.

Enrollment and the traits of raters: We found online raters using the International Society of Dermoscopy's email lists and social media pages. Evaluators had to sign up for the research and create an account with a username, email address, and password. The raters' ages, genders, countries of origin, occupations, and total years of dermatoscopy experience were also collected. To confirm that the self-reported experience matched the real abilities, each rater was required to do repeated screening tests. Simple domain-specific tasks, such as classifying melanoma from non-melanoma and seborrheic keratosis from other lesions, were used as part of the screening process. The initial interaction research, which screened several types of AI-based assistance, attracted 304 raters, whereas the extended interaction study and the telemedicine study both recruited 157 raters. Selecting 3 dermatology residents and 8 board-certified dermatologists as second-opinion raters between April 2022 and September 2022 since they have identified and handled more than two worrisome skin lesions in person. We recruited fourth-year medical students for the knowledge transfer research, and although only 202 of the 652 initially approached decided to participate, of those 189 answered more than half of the exam questions correctly.

3.3 Statistics

We limited the dataset to pictures that had at least three different assessments so that it may represent the aggregate judgments of a small human group. For every image, we compiled a total of 30 bootstraps, each consisting of three to five randomly picked ratings, and used the average of these scores as our collective forecast. The ties were cracked at random. Next, we determined the average accuracy for every image by averaging the fraction of accurate bootstrapped predictions. To combine human associations with HCI-DL-based predictions, we calculated an arithmetic mean of the sum of the human multiclass probabilities and the matching DL-based multiclass probabilities. To facilitate comparisons across raters with and without decision help, we computed geometric means for each user, and for diagnostic analyses, we averaged the findings for every image. Confident or non-confident responses were those that were either more or less rapid than the individual means, accordingly, based on the user's mean responding time across all interaction modalities.

We employed a specified threshold of 0.17 to indicate malignancy for the filtering technique in the telemedicine trial. Therefore we used it to determine whether or not a sample should be filtered. If a patient photographed a lesion more than once, all that was required was one picture that was over the cutoff for the lesion to be classified as "probably malignant," and the same was true on the patient level. Using a t-test for a single sample, we checked whether normally distributed continuous data showed any significant departures from zero to the collected information. The paired or unpaired t-test, as well as the Wilcoxon signed-rank test, were used, depending on the circumstances, to analyze the continuous data that was collected from each group. To analyze the differences in proportions, a chi-squared test was carried out. All of the reported P values were subjected to the Holm–Bonferroni40 correction for multiple screening, and a two-sided ($P < 0.05$) was seen as having statistically significant implications. R version 3.6.2 was used to conduct all of the analyses, while ggplot version 3.2.1 and ggalluvial version 0.11.1.1 were used to produce the plots.

4. Result and Discussion

Detecting SC using DL algorithms has shown encouraging results. To help in the early detection and diagnosis of SC, researchers have built deep learning models that can properly identify and organized skin lesions as early (non-cancerous) or malignant (cancerous). It is important to carefully validate and integrate deep learning models for SC detection into current diagnostic procedures before deploying them in clinical settings. It has been shown that HCI-DL algorithms can identify and classify SC lesions with high accuracy, with results that are on par with or even better than those achieved by dermatologists. Existing approaches, such as HCI-AI and conventional neural networks (CNN), are compared to our suggested method. The metric, which may include precision, sensitivity, and specificity.

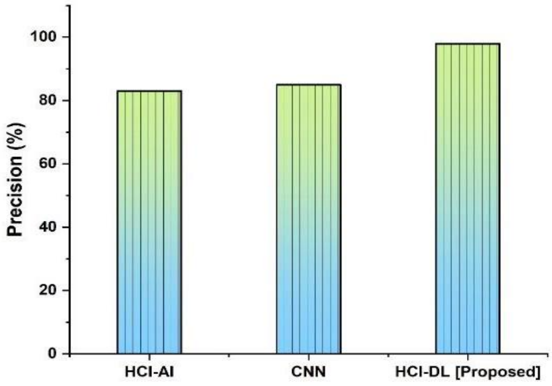

Fig. 2. Comparison of Precision

Precision is shown in figure 2 and is defined as the percentage of positive instances that were indeed cancerous. The accuracy of a cancer detection system is measured by how well it separates false positives from real ones. Here is an equation (1) for determining precision:

$$Precision = True\ Positives / (True\ Positives + False\ Positives) \quad (1)$$

The method accurately recognizes cancer cases without misclassifying non-cancerous cases as positive if the accuracy parameter is high. However, if the precision number is poor, it means that the system is misidentifying non-cancerous instances as malignant, which is not what we want. The suggested technique achieved a score of 98%, whereas the existing HCI-AI method achieved 83% and the existing CNN method achieved 85%. The suggested method's precision is superior to that of existing alternatives. The precision of both the proposed and current approaches are compared in Table 1.

Table 1. Precision

	Precision (%)
HCI-AI	83
CNN	85
HCI-DL [Proposed]	98

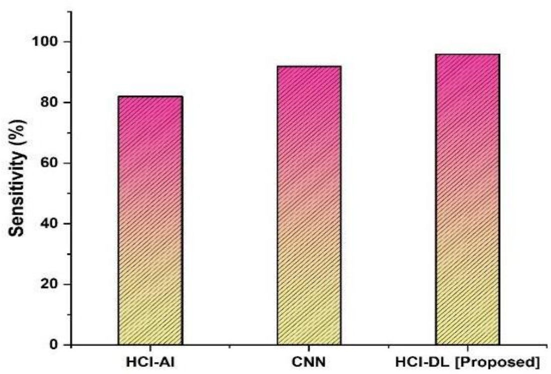

Fig. 3. Comparison of sensitivity

Figure 3 depicts the sensitivity, which is another significant statistic used to assess the efficacy of a cancer detection system and is also known as the recall or true positive rate. The level of sensitivity refers to how many genuine cancer cases can be positively diagnosed by the system. The sensitivity rate of the system in identifying instances of cancer is measured. The following equation (2) may be used to determine sensitivity:

$$Sensitivity = True\ Positives\ /\ (True\ Positives + False\ Negatives) \quad (2)$$

Sensitivity, like precision, has to be taken into account with other indicators of performance to give a whole assessment. The required degree of sensitivity might change based on the nature of the task at hand. Achieving high sensitivity in cancer detection is critical for ensuring that few cancer cases are overlooked, allowing for early identification and prompt treatment. When compared to the state-of-the-art methods of HCI-AI (82% accuracy) and CNN (92% accuracy), the suggested technique achieved a score of 96%. The suggested strategy outperforms other approaches due to its superior sensitivity. Table 2 shows the results of a comparison between the suggested and current approaches in terms of sensitivity.

Table 2. Sensitivity

	Sensitivity (%)
HCI-AI	82
CNN	92
HCI-DL [Proposed]	96

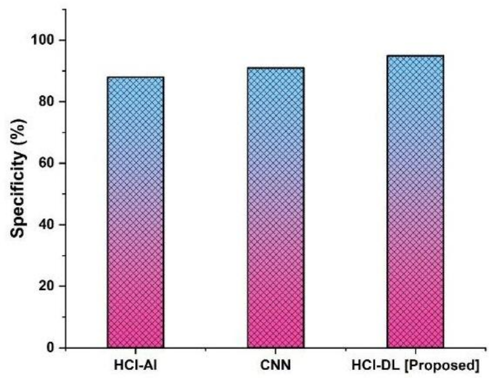

Fig. 4. Comparison of Specificity

Specificity is shown in figure 4; it is a measurement of how many non-cancer cases the system has accurately recognized as negative. The specificity measures how well the system can identify benign situations. The equation (3) for determining specificity is as follows:

$$Specificity = True\ Negatives\ /\ (True\ Negatives + False\ Positives) \quad (3)$$

In cancer detection, specificity is especially significant since it decreases the number of false positives and the number of people who get unneeded interventions or treatments. The suggested technique achieved 95% accuracy, whereas the state-of-the-art HCI-AI method achieved 88% accuracy and the CNN method achieved 91% accuracy. The suggested technique is more sensitive than current approaches. Table 3 displays a comparison of the proposed and current approach's specificity.

Table 3. Specificity

	Specificity (%)
HCI-AI	88
CNN	91
HCI-DL [Proposed]	95

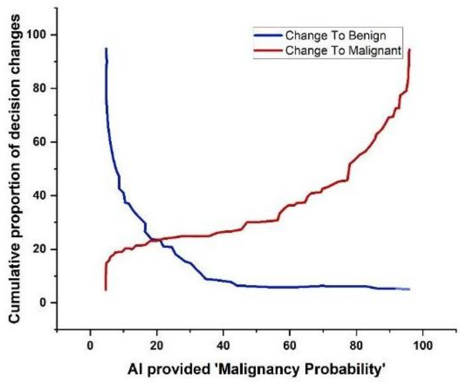

Fig. 5. Malignancy thresholds are arbitrarily set by raters

Figure 5 represents the average threshold for the AI-provided malignancy probability was not half percent but a quarter percent, whether shifting replies from benign to malignant or malignant to benign diagnosis. Researching and analyzing medical data from large groups of patients with a certain ailment, such as diagnostic test results, pathology reports, and treatment outcomes, is necessary for establishing malignancy thresholds. To better understand the relationship between illness features and patient outcomes, researchers collect and analyze this information.

5. Conclusion

This research exemplifies the obvious benefit of HCI assessment in guiding the creation of information systems like CDS. Images in this research were categorized using HCI-DL. If more people used this method, clinical software development, adoption, and everyone would gain from incorporation into practice. It is crucial to integrate quick user-centered iterative design with lab-based simulation for a brand-new CDS tool that is not yet in broad use. Incorporating human factors input early and iteratively into the design of clinical information systems has the potential to save costs while simultaneously improving user performance and usability. This is accomplished by addressing essential HCI and clinical workflow features before deployment. The DL method has to be enhanced, and since the GPU isn't operating quickly enough, processing images will take a long time. This research will upgrade its login process in the future, enabling users to see a history of prediction outcomes and moving the application to a Heroku-hosted public server. As well as collaborating with many hospitals to enable users to talk with cancer professionals directly for guidance on the information and outcomes of this initiative.

Reference

[1] Calisto, F.M., Santiago, C., Nunes, N. and Nascimento, J.C., 2021. Introduction of human-centric AI assistant to aid radiologists in multimodal breast image classification. International Journal of Human-Computer Studies, 150, p.102607.

[2] Sabol, P., Sinčák, P., Hartono, P., Kočan, P., Benetinová, Z., Blichárová, A., Verbóová, Ľ., Štammová, E., Sabolová-Fabianová, A. and Jašková, A., 2020. Explainable classifier for improving the accountability in decision-making for colorectal cancer diagnosis from histopathological images. Journal of biomedical informatics, 109, p.103523.

[3] Lyell, D., Coiera, E., Chen, J., Shah, P. and Magrabi, F., 2021. How machine learning is embedded to support clinician decision making: an analysis of FDA-approved medical devices. BMJ Health & Care Informatics, 28(1).

[4] Cai, C.J., Reif, E., Hegde, N., Hipp, J., Kim, B., Smilkov, D., Wattenberg, M., Viegas, F., Corrado, G.S., Stumpe, M.C. and Terry, M., 2019, May. Human-centered tools for coping with imperfect algorithms during medical decision-making. In Proceedings of the 2019 chi conference on human factors in computing systems (pp. 1-14).

[5] Fan, X. and Zhong, X., 2022. Artificial intelligence-based creative thinking skill analysis model using human–computer interaction in art design teaching. Computers and Electrical Engineering, 100, p.107957.

[6] Marcilly, R., 2019. Development of a Video Coding Scheme for Understanding Human-Computer Interaction and Clinical Decision Making. Context Sensitive Health Informatics: Sustainability in Dynamic Ecosystems, 265, p.80.

[7] Cai, C.J., Winter, S., Steiner, D., Wilcox, L. and Terry, M., 2019. "Hello AI": uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. Proceedings of the ACM on Human-computer Interaction, 3(CSCW), pp.1-24.

[8] Xu, W., 2019. Toward human-centered AI: a perspective from human-computer interaction. interactions, 26(4), pp.42-46.

[9] Bond, R.R., Mulvenna, M.D., Wan, H., Finlay, D.D., Wong, A., Koene, A., Brisk, R., Boger, J. and Adel, T., 2019, October. Human Centered Artificial Intelligence: Weaving UX into Algorithmic Decision Making. In RoCHI (pp. 2-9).

[10] Sardar, P., Abbott, J.D., Kundu, A., Aronow, H.D., Granada, J.F. and Giri, J., 2019. Impact of artificial intelligence on interventional cardiology: from decision-making aid to advanced interventional procedure assistance. Cardiovascular interventions, 12(14), pp.1293-1303.

[11] Rundo, L., Pirrone, R., Vitabile, S., Sala, E. and Gambino, O., 2020. Recent advances of HCI in decision-making tasks for optimized clinical workflows and precision medicine. Journal of biomedical informatics, 108, p.103479.

[12] Haesevoets, T., De Cremer, D., Dierckx, K. and Van Hiel, A., 2021. Human-machine collaboration in managerial decision making. Computers in Human Behavior, 119, p.106730.

[13] Tschandl, P., Rosendahl, C. and Kittler, H., 2018. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data, 5(1), pp.1-9.

[14] Alkatheiri, M.S., 2022. Artificial intelligence assisted improved human-computer interactions for computer systems. Computers and Electrical Engineering, 101, p.107950.

[15] Yuan, X., Zhang, L. and Zhao, S., 2023. DenseNet Convolutional Neural Network for Breast Cancer Diagnosis.

[16] Abinash Das. (2022). Effectiveness of Web-Based Decision Making to Deliver Decision-Support Information to Business Analyst using a 'Thin-Client. Technoarete Transactions on Advances in Data Science and Analytics (TTADSA), 1(1): 1-4. https://doi.org/10.36647/TTADSA/01.01.A001