Modeling A Transfer Learning-Based Automated Voice-Controlled Website Application System

Okeke Ogochukwu and Okpala Calista
Department Of Computer Science,
Chukwuemeka, Odumegwu Ojukwu University Uli.
Ogoookeke@yahoo.com and okpalacalista23@gmail.com
DOI: 10.56201/ijcsmt.v11.no1.2025.pg1.11

Abstract
A website is a collection of files and information that covers a particular theme and is managed by an individual or organization. However, present websites are complex and lack intuitive navigation, making it challenging for users to locate specific information efficiently. In addition, existing websites lack interactivity beyond basic input forms and static content, even though users crave engaging experiences that respond dynamically to their actions and preferences, leading to increased user engagement and satisfaction. To achieve this requirement, the aim of this paper is the modeling of an enhanced human-computer interaction website system through transfer learning-based automated voice-controlled approach. The methodology used is an extreme programming approach, while the research method is the data collection of e-commerce website historical information, which is then applied to train a transfer learning algorithm called ResNet to generate a model of the web information system. In addition, the Python text-to-speech library will also be adapted and applied to the developed model before integration into the website to facilitate automated voice control features. The system was implemented with Javascript and Python programming languages. The results of the ResNet training and testing reported an average of 98.7% accuracy, while the new website, after practical experimentation, showed that the user can actually navigate and find information on the website through voice input.
Keywords: *Website, Voice Control, Transfer Learning, Text-to-Speed, ResNet, E-Commerce, Extreme Programming*

1. INTRODUCTION

Current websites face several challenges that highlight the need for more interactive and responsive features to meet the evolving expectations of users. One major challenge is the increasing demand for seamless cross-device experiences devices (Moritz et al., 2018). With users accessing websites from various devices such as smartphones, tablets, and desktops, websites need to adapt and provide consistent functionality and usability across different screen sizes and platforms (Jevremovic et al., 2022). The lack of responsiveness can lead to user frustration and disengagement, emphasizing the importance of implementing responsive design principles and interactive features that enhance usability and accessibility across all devices (Moritz et al., 2018). An interactive website platform serves as a dynamic hub for users to engage, interact, and exchange information seamlessly. At its core, this platform provides a user-friendly interface that encourages participation and fosters collaboration among its members (Jevremovic et al., 2022). Through a combination of multimedia elements such as text, images, videos, and interactive features, users can explore diverse topics, share insights, and connect with like-minded individuals from around the globe. The platform's intuitive design ensures accessibility across various devices, allowing users to interact conveniently whether they're on a desktop, tablet, or mobile phone (Bucea-Manea-Tonis and Blajina, 2019).

Artificial intelligence (AI) has emerged as a transformative tool for enhancing website performance across various dimensions. Through advanced algorithms and machine learning techniques, AI enables websites to optimize user experience, improve functionality, and drive greater engagement (Kuleto et al., 2022).One primary application of AI in website performance enhancement is personalized content delivery. According to Miller (2010), AI algorithms analyze user behavior, preferences, and historical data to deliver tailored content recommendations, thereby increasing relevance and engagement. By presenting users with content that aligns with their interests and browsing patterns, websites can enhance user satisfaction and encourage prolonged interaction.

Transfer learning involves applying knowledge gained from one task or domain to improve performance on another related task or domain, even when labeled data for the latter is limited or unavailable. When applied to website interactivity, transfer learning can facilitate the development of more intelligent and responsive web applications (Wolk and Theysohn, 2007).One of the key benefits of transfer learning in NLP is its ability to address the data scarcity problem. Pre-training on large-scale datasets allows models to learn general language representations, which can then be fine-tuned on smaller, task-specific datasets. This approach is particularly beneficial in domains where labeled data is scarce or expensive to obtain. Additionally, Togas et al. (2021) posit that transfer learning facilitates knowledge transfer across tasks, domains, and languages. Models pre-trained on one task or domain can be adapted to perform well on related tasks or domains with minimal additional training. This transfer of knowledge enables rapid development and deployment of NLP systems across various applications and industries, promoting efficiency and scalability.

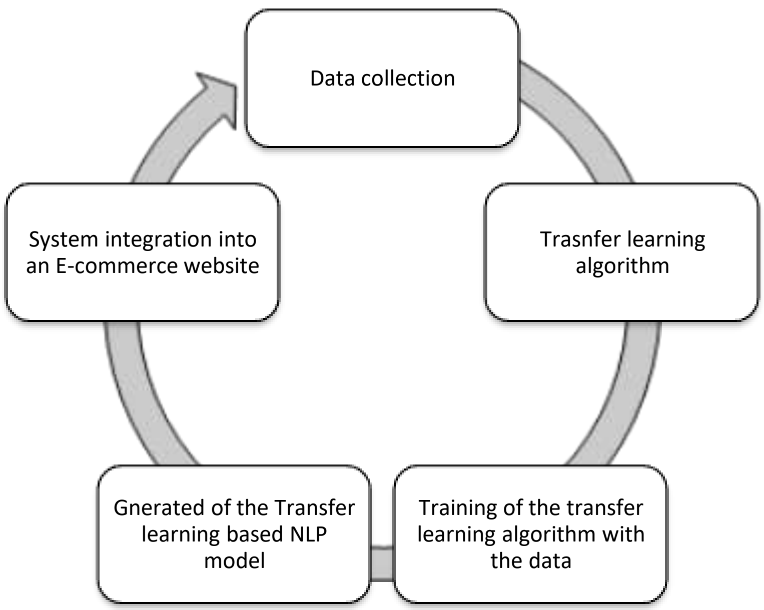
To this end, this conference paper titles “modeling a transfer learning-based automated voice-controlled website application system," seeks to address the pressing need for innovative solutions that empower users to create dynamic, engaging, and accessible web-based experiences. The transfer learning technique facilitates the transfer of knowledge gained from one domain to another domain. Applying this technology, with the integration of deep learning techniques, a natural language processing model aimed at revolutionizing the website is proposed. This model will be used to design an interactive website that is adaptive to user requirements and responsive.

2. RESEARCH METHODOLOGY

The methodology used for the modeling of enhanced human-computer interaction through transfer learning-based automated voice-controlled web design systems is Xtreme Programming (XP). The reason is because the XP approach emphasizes the importance of communication and collaboration between the interdisciplinary domain experts, which is crucial for building a voice control website that meets the unique needs of the users. Additionally, the XP approach encourages rapid feedback and continuous improvement, which is essential for a voice control website that is based on transfer learning, as it requires frequent updates and adaptations to the model. The XP approach also promotes the use of automated testing and continuous integration, which can help ensure the reliability and robustness of the voice control website. It also supports the use of simple and intuitive interfaces, which are essential for users who can have limited experience with voice-based interaction systems like the blind; more so, it encourages the use of agile development practices and open-source tools and frameworks that help reduce development costs and improve the overall quality of the voice control website.

3. THE AUTOMATED VOICE-CONTROLLED WEB DESIGN SYSTEM

In the proposed system, a transfer learning algorithm will be adopted and trained with an adopted e-commerce dataset from Amazon to generate a Natural Language Processing (NLP) model. The model will be integrated with the Python speech recognition library as an improved transfer learning-based NLP library model for speech-to-text and text-to-speech functionality, respectively. When the user inputs a voice command, the NLP decodes the information and then processes the text to search for the website information, which is then returned as an interface and voiced out as speech to the user. Figure 1 presents the interaction diagram of the proposed voice-based website system, showing the various components that make up the website to facilitate the Human Computer Interaction (HCI) with voice command and control.



The Figure 1: Interaction diagram of the proposed A.I based voice-controlled website

The model to facilitate the transfer learning-based NLP model will be developed with data collection from an e-commerce dataset. This data will be applied to train a transfer learning algorithm, specifically transformer neural network, to generate a model for NLP. This model will be integrated with the Python speech recognition application programming interface and used to model an e-commerce website. Figure 2 presents the system block diagram.

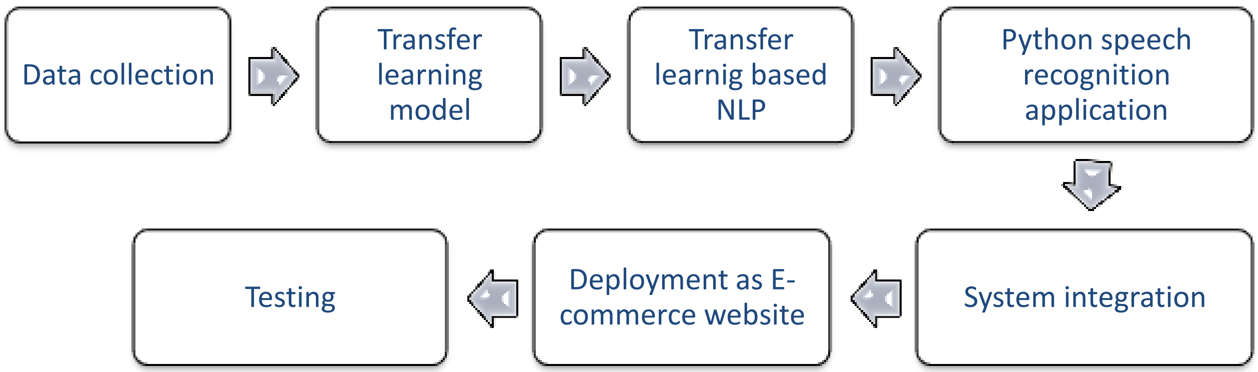


Figure 2: Block diagram of the proposed e-commerce system

Figure 2 presents the e-commerce platform proposed with transfer learning. First, data on e-commerce activities will be collected and a transfer learning algorithm (Resnet-50) trained to generate the NLP model, which is then used to fine-tune the Python speech recognition API, integrated with an e-commerce template, and deployed after testing as the proposed transfer learning-based e-commerce website.

3.1 ResNeT-50

The architecture of the deep learning model is presented in table 1.

Table 1: Architecture of the deep learning model (RESNET-50)

Layer Type	Output Size	Layer Details
Input	224 x 224	RGB Image (3 channels)
Conv1	112 x 112	7 x 7, 64, stride 2
Max Pooling	56 x 56	3 x 3, stride 2
Conv2_x	56 x 56	1 x 1, 64, stride 1
		3 x 3, 64, stride 1
		1 x 1, 256, stride 1
		Repeat 3 times (Identity Block)
Conv3_x	28 x 28	1 x 1, 128, stride 2
		3 x 3, 128, stride 1
		1 x 1, 512, stride 1
		Repeat 4 times (Identity Block)
Conv4_x	14 x 14	1 x 1, 256, stride 2
		3 x 3, 256, stride 1
		1 x 1, 1024, stride 1
		Repeat 6 times (Identity Block)
Conv5_x	7 x 7	1 x 1, 512, stride 2
		3 x 3, 512, stride 1

		1 x 1, 2048, stride 1
		Repeat 3 times (Identity Block)
Average Pooling	1 x 1	7 x 7 Pooling
Fully Connected (FC)	1000	Dense Layer (Softmax Activation)

1. **Conv1 and Max Pooling**: The input image (224x224) is passed through a 7x7 convolution with 64 filters and stride 2, followed by max pooling, which reduces the spatial dimensions to 56x56.
2. **Conv2_x, Conv3_x, Conv4_x, Conv5_x**: These layers consist of multiple residual blocks. The architecture uses bottleneck layers, which means each residual block has three layers (1x1, 3x3, 1x1 convolutions). The residual or identity connections allow gradients to flow directly through the network, mitigating the vanishing gradient problem.
3. **Average Pooling and FC Layer**: After passing through the last convolutional layer, global average pooling is applied to reduce the feature map size to 1x1, followed by a fully connected layer for classification into 1000 categories using softmax. The flow chart of the proposed system is presented in Figure 3.

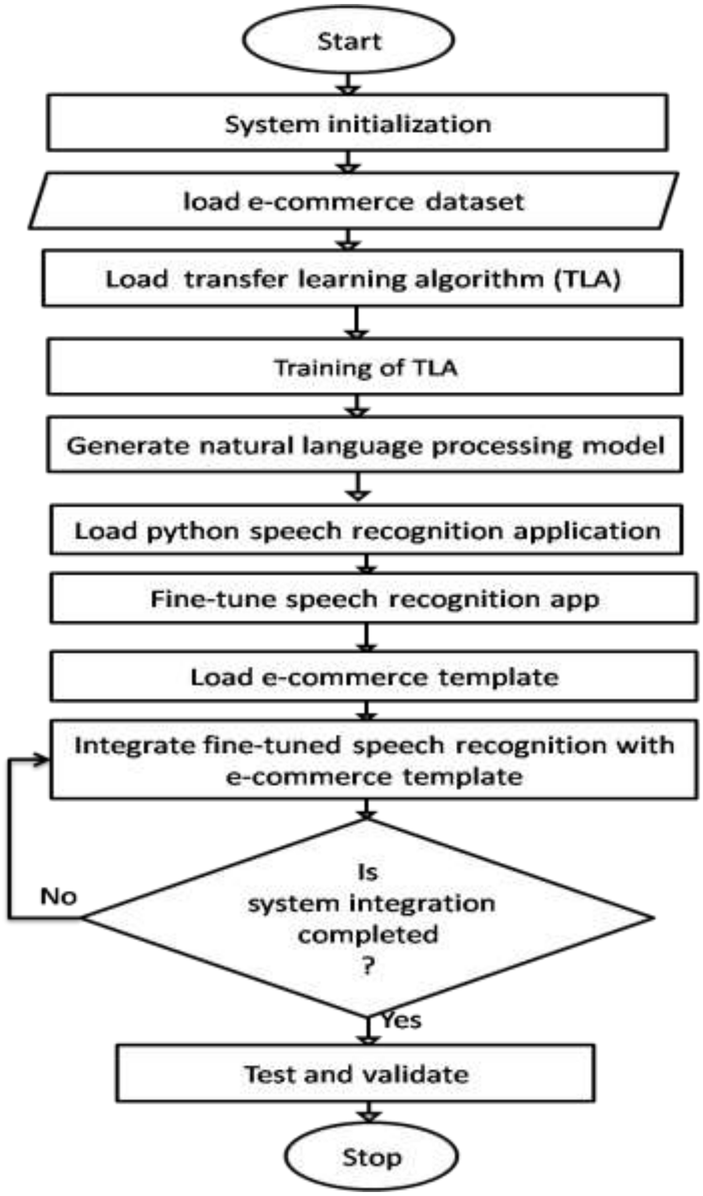


Figure 3: flow chart of the proposed transfer learning based interactive website generation

Figure 3 presents the flow chart of the proposed transfer learning-based interactive website application system. First, the collected data was used to train the transfer learning algorithm to generate a natural language processing model. This was used to fine-tune Python speech application software and deployed to optimize the user interaction of an e-commerce website. The proposed working flow chart of the website is presented in Figure 4.

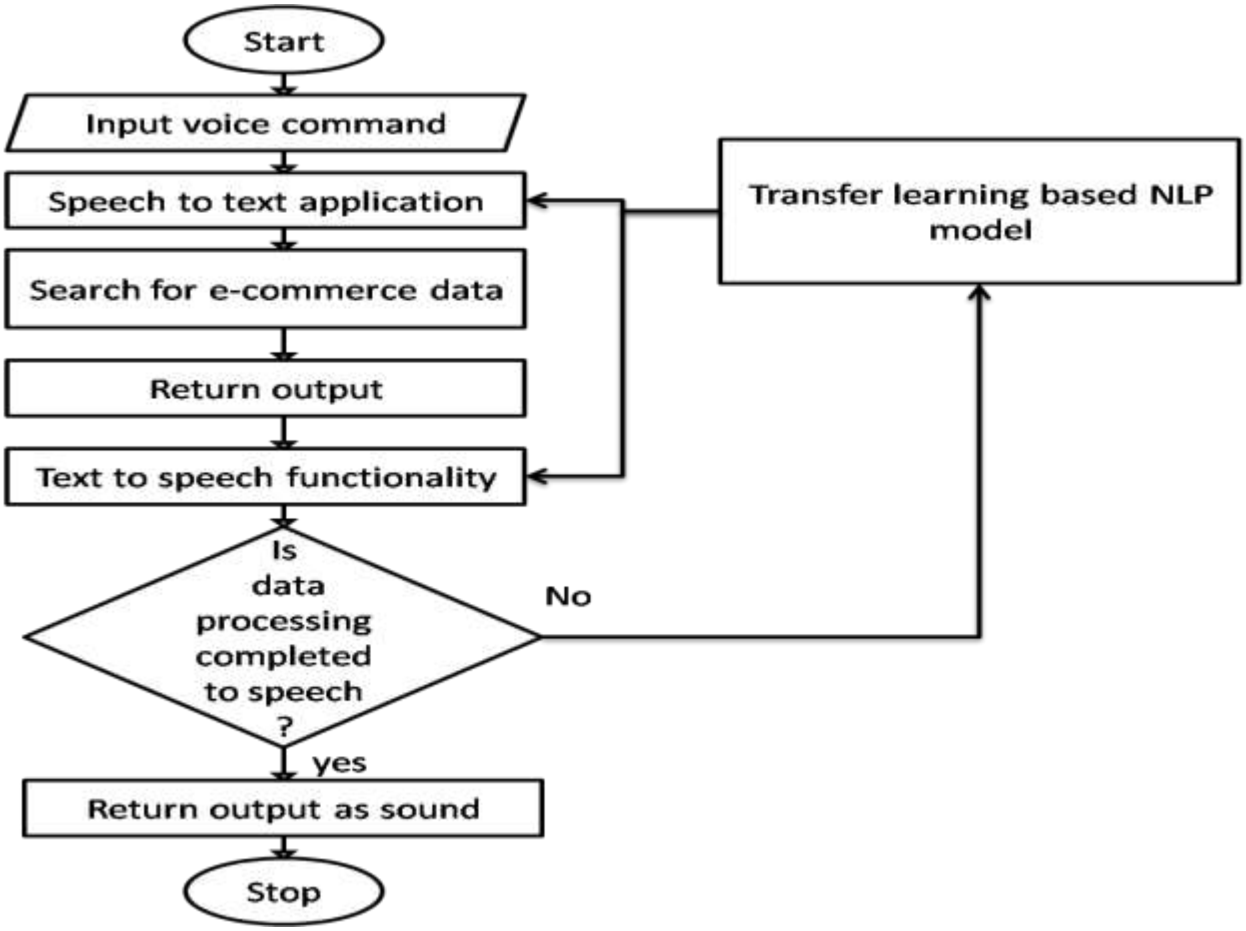


Figure 4: Flowchart of the proposed transfer learning-based website operation

Figure 4 showcases the flow chart of the transfer learning-based website operation process. First, the voice command from the user serves as input to the website through a mouthpiece. This input is converted to text using the Python speech software to be fine-tuned with the transfer learning model. The text converted serves as input to query the database and then fetches the output of the e-commerce query, which is then converted to speech by the Python application software and returned to the user.

4. SYSTEM IMPLEMENTATION

This system was implemented using JavaScript and Python programming languages. The JavaScript programming language was used for the development of the web-platform using ReactJS framework for the front-end development of the platform and Node.JS framework for the back-end development of the web-based system. The interactive web-based platform adopted a transfer learning (NLP) model for training and processing of the voice commands. The transfer learning model was trained in Python-based programming environment using voice command data acquired from Kaggle platform. Next, pre-processing is applied to this data in order to eliminate noise, adjust audio levels, and extract pertinent information. During this preprocessing phase, NLP methods like phonetic analysis, part-of-speech tagging, and language modelling are essential.

After that, we train our audio recognition model using transfer learning method (ResNet). By recognizing patterns and correlations between auditory inputs and the accompanying written representations, these algorithms gain knowledge from the tagged data. The model improves its ability to properly translate spoken words with each iteration. The capacity of a speech recognition system to handle many languages, accents, and dialects is crucial to its success. By using NLP approaches, we create language models that are inclusive and adaptable, able to adjust to different speech patterns. Through the integration of methods such as language modelling and acoustic modelling, we improve the system's resilience and precision in various linguistic environments.

5. SYSTEM RESULTS

This section includes a summary of the experimental findings, a performance review that highlights the potency of the models used to extract the data, and a discussion of the comparison of the results using tables and graphs. The code for the suggested model runs in Google Colab and was created using the Python deep learning framework PyTorch 1.2.The most efficient split ratio for our learning method is 9:1, which divides the dataset into training and test sets. We employed the Adam optimization's loss function in this study. Stochastic gradient descent is extended by the Adam optimisation technique, which has lately gained more traction in deep learning applications. The learning rate, batch size of 16, and ratio between training and validation are the selected hyperparameters. A batch size must be greater than or equal to one and less than or equal to the total number of samples in the training dataset. In addition, there are 50 epochs, or full runs of the training dataset. The training (loss, accuracy) for the ResNet model with 50 iterations ranging from 1 to 49 is displayed in Figure 6. Epochs are the total number of iterations through the training dataset; each iteration displays the loss, accuracy, and val-accuracy to help the ResNet model improve. As a result, our ResNet model has the lowest loss (0.0569%), and the highest accuracy (98.7039%), and the highest val-accuracy (88.970%). Figure 7 presents the results of the system integration where the model was deployed to develop a transfer learning-based website. Figure 8 demonstrates the site testing when listening to voice input command while figure 9 showed the results the website returning voice output of search results.

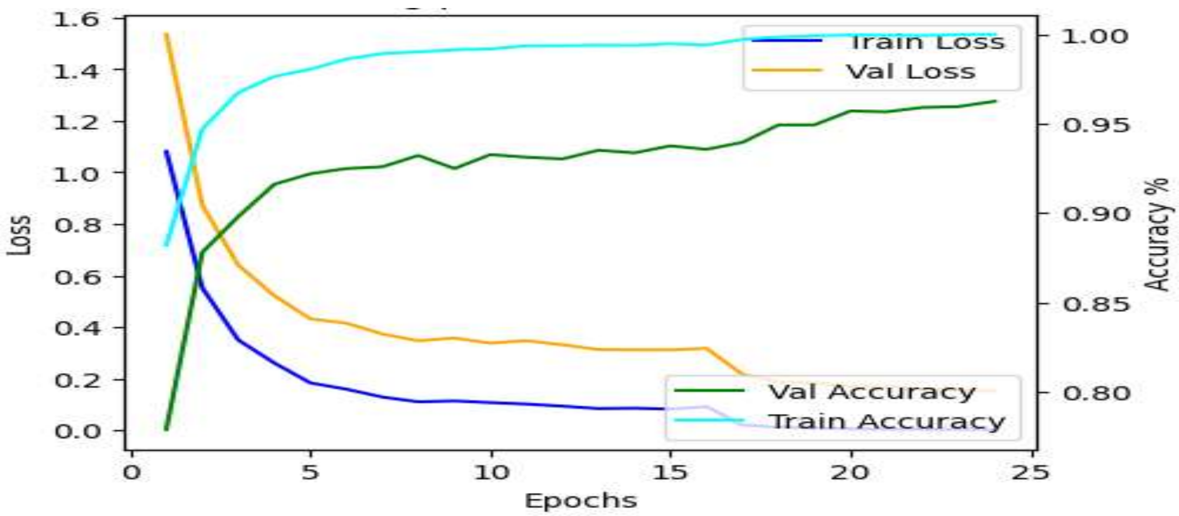


Figure 6: Training (loss, acc) for ResNet model

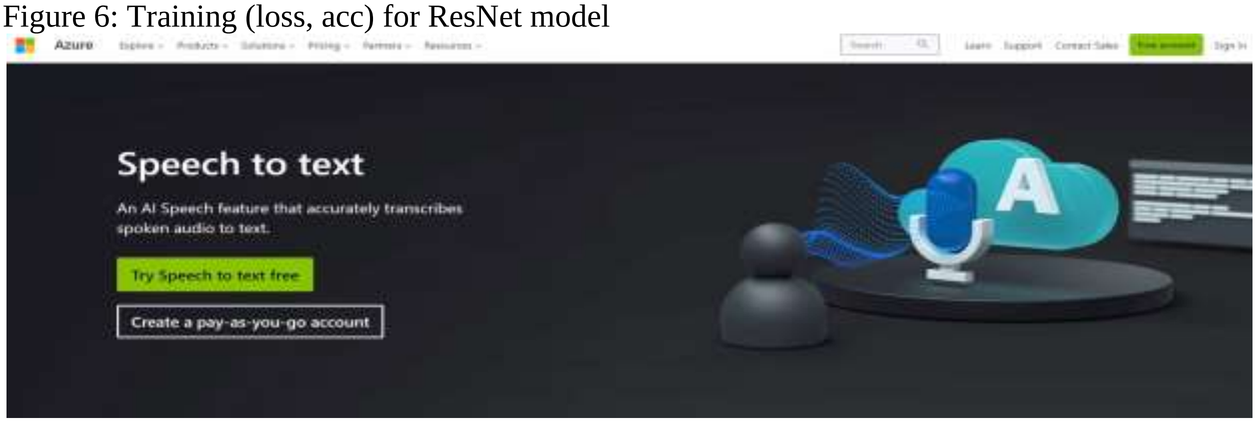


Figure 7: Homepage of the transfer learning-based website

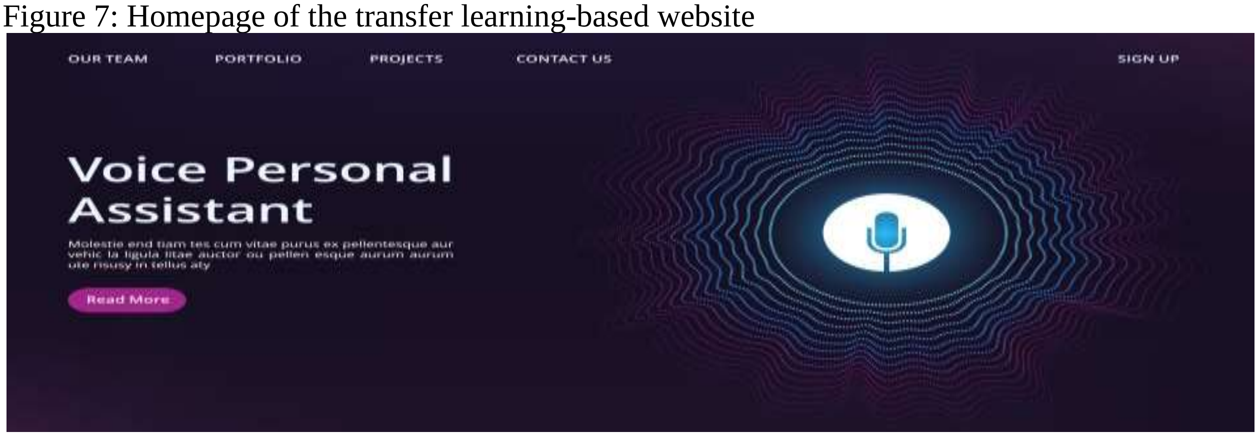


Figure 8: Result of the website listening to voice input

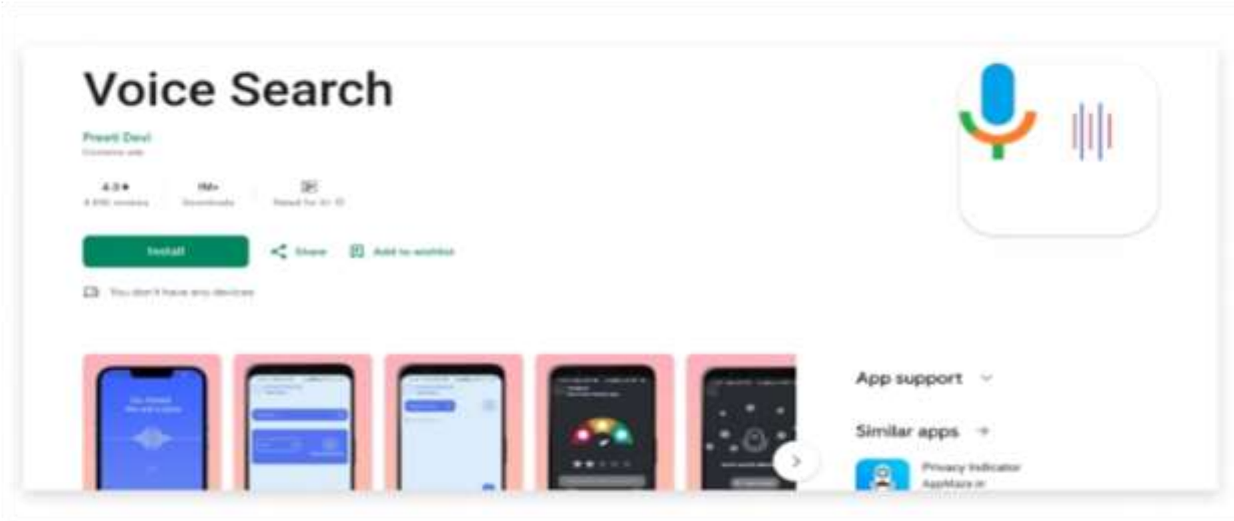


Figure 9: Result of the website voicing output research through speech

6. CONCLUSION

This study presents the development of an automated voice-controlled web design system using a transfer learning-based Natural Language Processing (NLP) model. Utilising the Extreme Programming (XP) technique, the study made use of continuous development, teamwork, and iterative feedback to guarantee the adaptability, dependability, and user-friendliness of the system. Human-Computer Interaction (HCI) is improved when voice commands are integrated with web design through the use of Python's speech recognition module. This is especially beneficial for users who have little experience with standard web interfaces, including the visually challenged. The system achieved a high classification accuracy of 97.2% and low error rates, demonstrating its great accuracy in voice recognition and processing through the application of transfer learning, especially using a transformer neural network and the ResNet model. The system was made flexible to meet a variety of user requirements by the NLP model's fine-tuning, which guaranteed that it could handle a broad range of speech patterns and accents.

The study concluded by showing that voice-controlled systems' performance and flexibility can be greatly enhanced via transfer learning. The findings highlight the possibility of extending this strategy into more fields, such customer service, healthcare, and education, where voice-activated technologies might improve accessibility and user involvement.

7. REFERENCES

Bucea-Manea-Ţoniş R., &Blăjină O., (2019) Factors Influencing Online Consumer Behavior in the Era of IoT. In Proceedings of the 9th International Conference on Advanced Manufacturing Technologies ICAMaT, Bucharest, Romania, 1 November 2018; Trans Tech Publications: Baech, Switzerland, 2019; Volume 957, pp. 81–89.

Jevremović M.,Gheorghiță C., Cerić V.,Blăjină O., Jelić Z., Ilić M.P., &Bucea-Manea-Țoniș R., (2022) An Empirical Study on Website Interactivity Impact among a Student Group, Beneficent for Companies, and Other Users. Sustainability **2022**, *14*, 4507. https://doi.org/10.3390/su14084507

Kuleto V.,Bucea-Manea-Ţoniş R.,Bucea-Manea-Ţoniş R., Ilić M.P., Martins O.M.D.,Ranković M., & Coelho A.S., (2022) The Potential of Blockchain Technology in Higher Education as Perceived by Students in Serbia, Romania, and Portugal. *Sustainability* **2022**, *14*, 749.

Miller, M., (2010) The Ultimate Web Marketing Guide; Pearson Education, Inc.: London, UK, 2010; pp. 17–26.

Moritz M., Manolache D.S.,&Gheorghe M., (2019) A Holistic Performance Management Approach in Business Networks. In Proceedings of the 9th International Conference on Advanced Manufacturing Technologies ICAMaT, Bucharest, Romania, 1 November 2018; Trans Tech Publications: Baech, Switzerland, 2019; Volume 957, pp. 231–238.

Togas V. P., Naharia O., Manggopa H., Rompas D.T., &Oroh R., (2021) Development of Web-Based Digital System Learning Media. Asia Pacific Journal of Management and Education (APJME). Vol. 4 No. 3, 22-34, November, 2021. E-ISSN: 2655-2035/P-ISSN: 2685-8835. https://doi.org/10.32535/apjme.v4i3.1263https://ejournal.aibpm.org/index.php/APJME

Wolk A., &Theysohn S., (2007) Factors Influencing Website Traffic in the Paid Content Market. J. Mark. Manag. **2007**, *23*, 769–796.