DOI: 10.24996/ijs.2024.65.4.38

ISSN: 0067-2904

Islamic QA with Chatbot System Using Convolutional Neural Network

Ratih N.E. Anggraini*, Dara Tursina, Riyanarto Sarno
Informatics Department, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Received: 25/10/2022 Accepted: 17/4/2023 Published: 30/4/2024

Abstract

Many questions and answers about Islamic law are scattered on the internet and have been explained repeatedly by various sites. One solution is presented by the website www.piss-ktb.com, which creates a web-based source of information in the form of Frequently Asked Questions (FAQ). However, web-based FAQs have a weakness: users still have to browse through the available questions one by one according to the questions they want to know the answers to. Browsing through thousands of FAQs is inefficient and exhausting.

Thus, a chatbot system can become a better alternative to the FAQ website. Still, chatbots are difficult to use because most of their conversations are hard to understand. A single character error will cause the system to misunderstand its meaning. In reality, users expect a chatbot that can understand everyday language. Thus, it is necessary to develop a chatbot system that can understand various common sentence combinations in everyday language and understand the meaning of words. In addition, it should be able to predict answers automatically to various kinds of questions and requests, even though the initial training data is relatively low. Therefore, this study aims to develop a system that can provide answers automatically based on user commands in natural language using Global Vectors for Word Representations (GloVe), Convolutional Neural Networks (CNN), and Transfer Learning techniques. The result shows that the use of transfer learning and the Nadam optimizer can improve the system's performance.

Keywords: Chatbot, Convolutional Neural Network, GloVe, IslamicQA, Nadam optimizer, Transfer Learning

1. Introduction

The ability of chatbots to provide a personal touch with their interlocutor is considered to provide a new nuance in the world of human-computer interaction (human-computer interface). This concept is called a "conversational interface." ELIZA is a chatbot that was first created in 1966 [1]. The simple goal at that time was to find out whether ELIZA could deceive humans so that humans felt they were communicating with humans. Long story short, this research continues and produces another chatbot called Alice [2]. ALICE's ability to communicate like humans used to have been successfully realized by Loebner (chatbots that contest how long chatbots can fool a human) in 2000, 2001, and 2004.

Another remarkable example is speech-based chatbots, such as Siri [3], Alexa [4], Cortana [5], and Google Assistant [6]. These types of chatbots serve as virtual assistants [7] that can communicate with voice commands to create reminders, schedule appointments, find nearby places, and more. Chatbots have become an alternative to replacing jobs involving conversations with users, where this work is simple and repetitive.

*Email: ratih_nea@if.its.ac.id

On the other hand, there is a lot of information on the web that can be a source of information, one of which is through frequently asked questions (FAQ) websites [8]. These websites have been used by millennials to find desired information regarding Islamic law [9]. However, this method is less efficient and time-consuming since they have to search the article one by one. The chatbot is the solution to this problem because questions regarding Islamic law can be directly answered by this system. Unfortunately, ordinary chatbots also have a drawback: they can only answer questions based on certain keywords and cannot understand conversations.

This problem can be overcome with a chatbot system that can understand various sentence combinations used in everyday language. The chatbot system must understand the meaning of words and automatically predict answers to various user questions and requests, even though the initial training data is relatively small [10]. We need a chatbot to understand human language by leveraging NLP machine learning.

In previous research, a chatbot with an Adam optimizer has been built [11]. Adam's method has been proven to work better than other adaptive learning methods such as AdaGrad, AdaDelta, and RMSProp. Thus, it becomes the algorithm often used to train neural networks. Adam One is an important algorithm in the learning process for the parameters of neural networks: gradient descent. This algorithm plays a role in updating the weights of each neuron.

In this research, however, we used Nadam as the gradient descent algorithm of the optimizer. Nadam is a combination of Adam and Momentum. In some cases, Nadam is considered to have the ability to surpass Adam because Nesterov's momentum is better than conventional momentum. It is evident from the evaluation results that using the Nadam optimizer is better than the Adam optimizer [12]. The advantage of this system is that, besides being able to handle typos, it also utilizes word embedding so that the system can handle input words that have existing words.

To investigate the abilities of Adam and Nadam optimizers, this work will combine CNN, the transfer learning method, and both optimizers in developing a chatbot system for Islamic QA. The state of the art and related works are explained in Section 2. Section 3 describes the proposed method and evaluation scenarios. The result and analysis are presented in Section 4. Finally, we conclude the research in Section 5.

2. Related Works

2.1 Transfer Learning

Transfer learning is a method of using knowledge from one task to perform another related or similar task [13]. Transfer learning aims to improve the learning process in the target task by taking advantage of the knowledge from the source task. This method is mostly used in inductive learning tasks where the model will use training data to perform a prediction on a case such as classification.

Research conducted by Pathak, Yadunath, et al. [14] employed deep transfer learning to classify the chest CT images in COVID-19 patients into COVID+ and COVID- classes. The result shows that transfer learning can increase the accuracy of COVID classification in chest CT images. Transfer learning combined with a convolutional neural network (CNN) has also been used to detect brain tumors from MR images [15]. While in the NLP, transfer learning was commonly used for text classification, document ranking, sentiment analysis, and question-answering tasks [16].

2.2 Convolutional Neural Network

A convolutional neural network is an improvement over an artificial neural network (ANN), in which the neurons can perform self-optimization through learning [17]. A basic CNN consists of four layers: the input layer, the convolutional layer, the pooling layer, and the fully connected layer. CNN has been used in many applications, from image classification to object recognition [18], from emotional analysis to speech recognition [19].

2.3 Recent Works

One of the recent studies regarding chatbots was conducted by Jiao using RASA NLU and neural networks [20]. RASA NLU is an open-source natural language understanding chatbot library. The performance of RASA NLU was better compared to common neural networks on a simple sentence and entity.

Another chatbot for health care was built using stochastic gradient descent (SGD) and the Adam optimizer [21]. Adam shows greater accuracy than SGD and ADAGRAD. The use of an optimizer indicates an increase in performance for the chatbot system. Thus, in this work, Adam and Nadam optimizers are implemented and their performances are compared.

3. Proposed Method

The method proposed in this study was to compare the results of CNN with and without transfer learning using the Nadam optimization algorithm. The transfer learning method required data sources as input into CNN and produced models. The proposed method is shown in Figure 1.

In this method, our documents will go through text preprocessing, which consists of stopword removal, tokenization, and stemming. The result of preprocessing is preprocessed text, which then undergoes sentence mapping using GloVe word embedding, producing a sentence matrix. Subsequently, stratified shuffle splitting will be conducted, and the output will be used for data training with CNN Learning. The CNN model that has been developed will be used to predict data testing.

3.1. Word Embedding Model

Word embedding is the method used to search for similar words based on their accompanying words. The more a word is mentioned together, the more it is seen. The word embedding used in this case study is “global vectors” (GloVe) [22]. Global Vectors for Word Representations (Glove) is an unsupervised learning algorithm for obtaining vector representations of words (word embeddings). The GloVe model training process is carried out by involving all statistical information from the corpus by forming a word co-occurrence matrix [23]. The GloVe training process is more efficient than other techniques because GloVe only involves training on non-zero matrix elements. The output of GloVe is a word vector representing each word’s meaning (word embeddings).

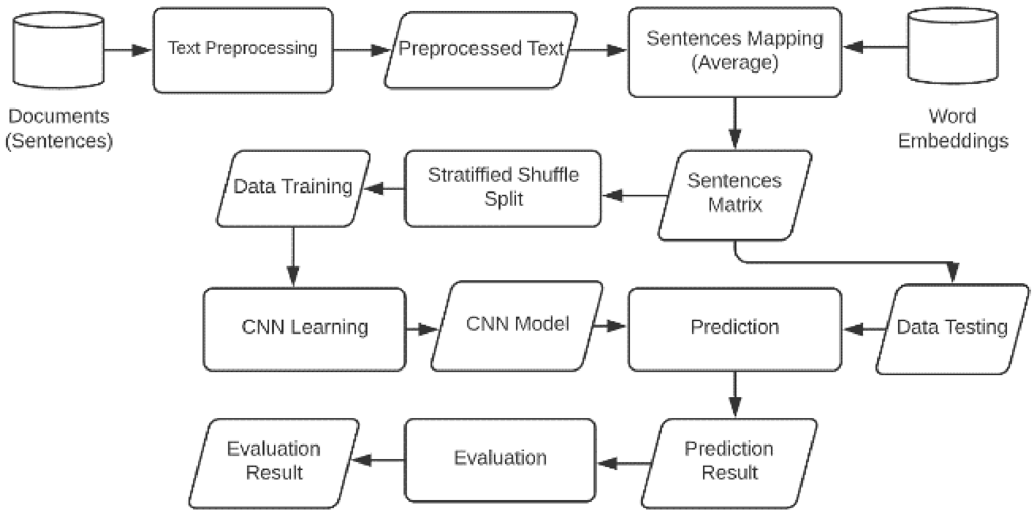

Figure 1: Proposed Approach

To get Word embeds, GloVe needs a large corpus as input. What is meant by "large" here is a corpus with a very large number of tokens (which can be tens of millions or even billions). If the corpus used is not large enough, then GloVe will fail to provide a representation of the meaning of each word. The GloVe word embedding process is shown in Figure 2.

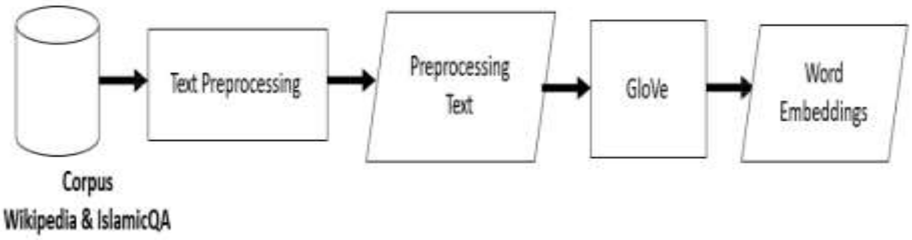

Figure 2: Glove Word Embedding Process

3.1.1. Sentence Mapping

Convolutional neural networks (CNN) were used four times. That is, in the source task and the target task, CNN transferred to Transfer Learning Scenario 1, and CNN transferred to Transfer Learning Scenario 2.

After each document goes through the text preprocessing stage, these documents will go through a process called "sentence mapping," where the vector representation of the sentence will be searched. The vector representation of the document, or so-called sentence embedding, is found by averaging each word embedding that composes the document. Each word that composes a document has word embeds stored in the database. If a word in the document is not in the database (does not have Word Embeddings), then the word is not considered to be in the document [23].

When all documents have obtained their vector representation, a matrix is formed that stacks all the sentence embeddings of each document. Each label class is also transformed into a one-hot vector. One hot vector is a vector that contains the number 0 as much as the number of existing classes and contains the number 1, which indicates the class for a document. This collection of one-hot vectors is also stacked into a matrix. These two matrices are input for CNN [24].

3.1.2. CNN Learning

Before doing the learning process with CNN, it is important to determine the CNN architecture itself. Architectures that match the dataset will provide the best performance. To get the best architecture, various experiments were carried out using the combination of parameters on the CNN. In the training/learning process using neural networks, the best model is often not obtained at the last epoch. Therefore, the learning process with CNN will involve checkpoints and early stopping. The architecture can be seen in Figure 3.

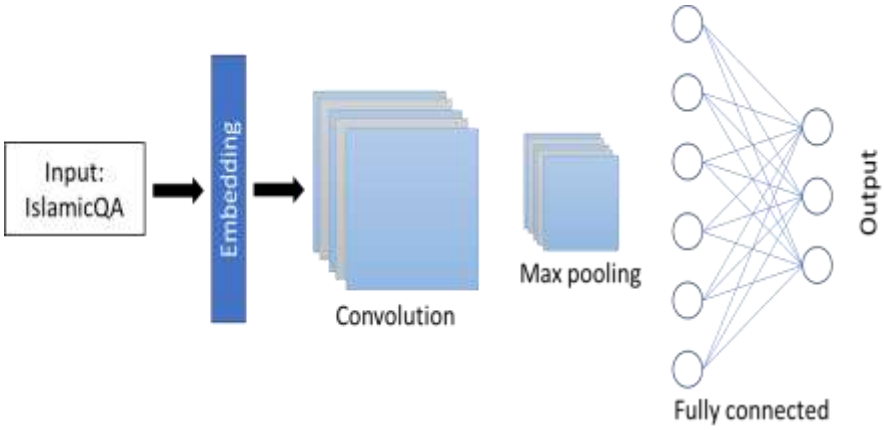

Figure 3: CNN Architecture on IslamicQA chatbot

Checkpoint is a technique for storing CNN modeling every time the loss value drops by a certain difference. That way, when the loss value is constant or tends to increase, the CNN model that managed to reach the lowest loss value will be stored.

Early stopping is a technique to stop the CNN learning process when the loss value does not show a significant decrease in a certain number of epochs. This method is used because it can optimize the number of epochs as much as possible, but it saves more time by stopping CNN training when it does not show an increase in learning. Gradient Descent is one of the important algorithms in the learning process of the neural network parameters, and this algorithm plays a role in updating the weights of each neuron [25].

Implementation of CNN was carried out with three scenarios. First, without transfer learning, Secondly, transfer learning scenario 1 uses a model trained with the IslamicQA data source. Thirdly, transfer learning scenario 2, with a model trained with the IslamicQA dataset and a model trained with the PISS-KTB dataset. The architecture of the three scenarios is equated by specifying ten convolution layers, and the dropout value is 0.1. The number of epochs set was 2000 with 10 epochs of patience, while Transfer Learning Scenario 2 was set to 1000 epochs with 10 epochs of patience. In the CNN model, the adaptive gradient descent algorithm was Nesterov-accelerated Adaptive Moment Estimation (Nadam) with a learning rate of 2x10-3 and 4x10-3. Nadam is a combination of Adam and Momentum [26]. In several cases, Nadam is considered to have the ability to surpass Adam because Nesterov's momentum is better than conventional momentum. The GloVe and CNN trials are implementations of the IslamQA dataset. The test in this section aims to see the ability of CNN with GloVe input without transfer learning.

3.2 Evaluation Scenarios

3.2.1. Dataset

This study aims to compare the results of CNN with those without transfer learning. In the transfer learning method, a data source is required as input in CNN to produce a model. The knowledge from this model is then transferred (partially retrieved and rebuilt) to be retrained with the target dataset. The requirements of the source dataset must also be identical to those of the target data. The source data was collected from https://islamqa.info/id, and the target data was from www.piss-ktb.com [27].

To measure the model's performance, the Sentence Embeddings set matrix and its labels are separated into two parts: training data and testing data. Training data is data that becomes CNN's material for learning. After the learning process has been conducted, the best CNN model will be reloaded to predict from each test data document. The label of each test set in the form of a one-hot vector matrix will be the ground truth of the CNN prediction results.
The dataset that has been preprocessed with embedding as a feature is then split into training data and testing data. The source dataset's distribution is divided into 10% test data and 90% training data. For the target dataset, the distribution is divided into 30% test data and 70% training data.

3.2.2. Performance Measurement

The prediction was carried out to measure the ability of the best CNN model based on previous training results. The CNN model is given input from the testing data. For each incoming sentence embedded, the output layer of CNN was the probability of each class with the softmax formula. The neuron with the highest output probability value is the prediction result from CNN.

The evaluation was conducted after each sentence embedding class was predicted. This step measured the performance of the model. At this stage, the prediction results of the model are compared with the ground truth. To make it easier, a confusion matrix was made to get the values of accuracy, recall, precision, and F-measure from each evaluation model.

3.3. Chatbot Application Design

Once we have the best-performing model from Transfer Learning, the model will be used to predict the types of sentences entered in the chatbot system. In this research, the LINE Messenger API (application program interface) was used to interact with the chatbot through the LINE application. Figure 3 explains how the chatbot application architecture is implemented so that answers to commands or questions from user chats are predicted and answered by bots.

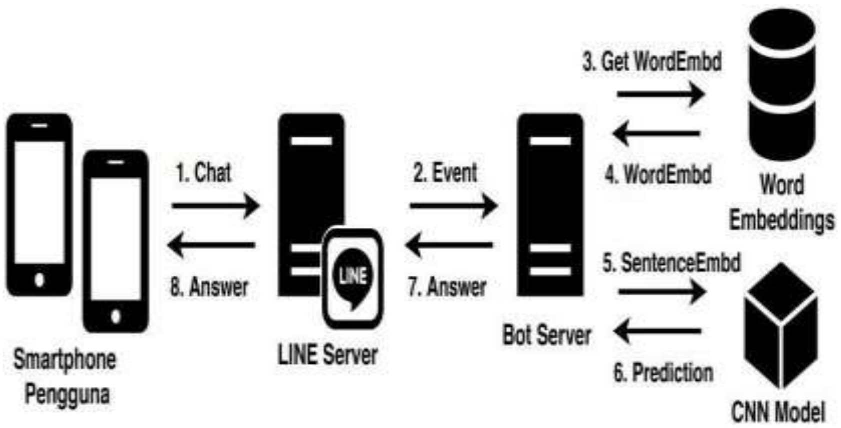

Figure 4: The communication flow in IslamicQA chatbot application

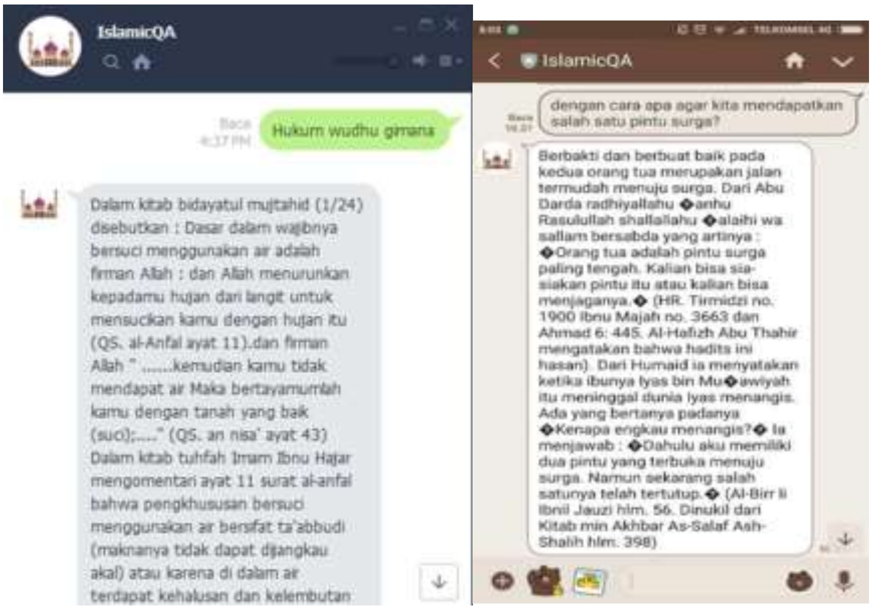

Figure 5: The interfaces of IslamicQA Line chatbot with different skins and questions

Chats were sent to the LINE server. Subsequently, with the webhook, the LINE server sent the chat to the bot server as an event. The event contained various kinds of sender data, including the sender's name, chat content, chat time, and others. The interface of the Line IslamicQA chatbot is shown in Figure 4.

The get_prediction function will accept a document from the incoming chat. This document is first converted into sentence embeddings using the vector function. Subsequently, CNN predicts the class by returning each class's probability value, which is stored in the prediction variable. In the prediction variable, the probability value of the first class occupies the first index, and the second-class probability value occupies the second index.

4. Result and Analysis

The Python programming language is implemented to propose the method. Python handles natural language processing (NLP) and machine learning tasks.

4.1. Chatbot Application Design

One of the reasons for using GloVe in this study is its ability to handle different words with similar meanings. In this experiment, three words were sampled to see if they had a close relationship with the use of other words that appeared in the dataset. It is shown in Table 1

Table 1: The resulting example of the 10 closest words

Indonesian Query	Similarity score	Indonesian Query	Similarity score	Indonesian Query	Similarity score
zakat		*Wudhu*		*Qurban*	
Infaq	0.767	Bilas	0.703	Gourlay	0.702
Fithrah	0.720	mandinya	0.699	Akikah	0.638
Halal	0.700	berwudhu	0.673	Menyembelih	0.629
Infak	0.684	Wudlu	0.671	Penyembelihan	0.622
Pemeluknya	0.673	Ember	0.651	Ekorjumlah	0.620
Shadaqah	0.671	Mandi	0.629	Odoru	0.617
Shodaqoh	0.648	berwudlu	0.626	Diharamkan	0.604
Wakaf	0.637	berwudu	0.619	Zakat	0.593
Umroh	0.630	Toilet	0.619	Ruminansia	0.583
Nisab	0.628	Wudu	0.599	Kurban	0.582

Based on this experiment, it can be concluded that GloVe has succeeded in providing a good representation of the meaning of each word. Words that have the same meaning have adjacent vectors. It was very important because chatbots are implemented in the field and deal with various types of users with unexpected uses of words, and GloVe can solve this problem well.

4.2. Performance Quality

The experiment was conducted with different scenarios mentioned in the previous sections. Table 2 describes the results of using the Adam Optimizer with and without the transfer learning process. The result shows that transfer learning is able to increase the performance of the system. Subsequently, Nadam Optimizer was used, and the results exceeded the performance of Adam Optimizer, as shown in Table 3.

It can be concluded that the combination of CNN and GloVe gave very good results (all metric values showed values above 94%). Preceded by GloVe's results, which have provided a good representation of the meaning of words, CNN has become much easier to classify. GloVe and CNN have proven successful in classifying sentence combinations and can predict other sentence combinations accurately.

The combination of GloVe, CNN, and transfer learning managed to cope with the scant data. The combination of the models of these three algorithms was able to achieve 91.17% accuracy, with the values of precision, recall, and F-measure being 94.29%, 91.17%, and 90.90% on the predicted data, respectively. Therefore, it can be concluded that the chatbot system developed in this research could understand the meaning of words and predict answers to various combinations of sentences with relatively little initial training data.

Table 2: The experiment result of the Adam Optimizer

Optimizer Adam	Accuracy (%)	Precision (%)	Recall (%)	F-1 Measure (%)
Without Transfer Learning	88.97	93.86	88.97	89.18
Transfer Learning Scenario 1	94.11	96.55	94.11	93.79
Transfer Learning Scenario 2	93.38	95.58	93.38	93.51

Table 3: The experiment result of the Nadam Optimizer

Optimizer Nadam	Accuracy (%)	Precision (%)	Recall (%)	F-1 Measure (%)
Without Transfer Learning	91.17	94.29	91.17	90.90
Transfer Learning Scenario 1	95.58	97.54	95.58	95.60
Transfer Learning Scenario 2	94.11	96.29	94.11	94.08

5. Conclusions

In this research, the CNN implementation used Nadam, the gradient descent algorithm, as the optimizer. Nadam is a combination of Adam and Momentum. In several cases, Nadam is considered to have the ability to surpass Adam because Nesterov's momentum is better than conventional momentum. The evaluation results prove that using the Nadam Optimizer algorithm is better than the Adam Optimizer algorithm, where accuracy, precision, recall, and F-measure results are all increasingly high.

For future research, we will combine GloVe's word embeddings with TF-IDF. TF-IDF is a very good method of determining the relevance of a word in a document. The unique word in a document should be prominently represented in the document. Thus, it can be assumed that multiplying the TF-IDFT value of a word by its word embeddings will produce a more accurate representation, making it easier for CNN to classify. Moreover, we can add stopwords in preprocessing for each document to be classified. By eliminating stopwords from the document, it is assumed that only highly relevant words will be classified.

6. Disclosure and conflict of interest

Conflict of Interest: The authors declare that they have no conflicts of interest.

References

[1] J. Weizenbaum, "ELIZA-A computer program for the study of natural language communication between man and machine," in *Communication of ACM*, vol. 9, no. 1, pp. 36–45, 1966.

[2] H. Shah. "ALICE: an ACE in Digitaland," *TripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, vol. 4, no. 2, pp 284-292, 2006

[3] Hasan, Rajibul, Riad Shams, and M. Rahman, "Consumer trust and perceived risk for voice-controlled artificial intelligence: The case of Siri," *Journal of Business Research*, vol. 131, pp 591-597, 2021.

[4] I. Lopatovska, et al., "Talk to me: Exploring user interactions with the Amazon Alexa," *Journal of Librarianship and Information Science*, vol. 51, no. 4, pp: 984-997, 2019.

[5] R. Mehrotra, et al., "Hey Cortana! Exploring the use cases of a Desktop based Digital Assistant," *SIGIR 1st international workshop on conversational approaches to information retrieval (CAIR'17)*, Vol. 4, 2017.

[6] N. Shakhovska, O. Basystiuk, and K. Shakhovska, "Development of the Speech-to-Text Chatbot Interface Based on Google API," *MoMLeT*, 2019.

[7] M.B. Hoy, "Alexa, Siri, Cortana, and more: an introduction to voice assistants," *Medical reference services quarterly*, vol. 37, no. 1, pp 81-88, 2018.

[8] W. Sakata et al., "FAQ retrieval using query-question similarity and BERT-based query-answer relevance," *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2019.

[9] S.K. Hamed and M.J. Ab Aziz, "A Question Answering System on Holy Quran Translation Based on Question Expansion Technique and Neural Network Classification," *J. Comput. Sci.*, vol. 12, no. 3,pp 169-177, 2016.

[10] B. Setiaji and F.W. Wibowo, "Chatbot using a knowledge in database: human-to-machine conversation modeling," *2016 7th international conference on intelligent systems, modelling and simulation (ISMS)*, IEEE, 2016.

[11] N. Bhartiya et al., "Artificial neural network based university chatbot system," *2019 IEEE Bombay Section Signature Conference (IBSSC)*, IEEE, 2019.

[12] M.N. Halgamuge, E. Daminda, and A. Nirmalathas, "Best optimizer selection for predicting bushfire occurrences using deep learning," *Natural Hazards*, vol. 103, no. 1, pp 845-860, 2020.

[13] L. Torrey and J. Shavlik, "Transfer learning: Handbook of research on machine learning applications and trends: algorithms, methods, and techniques," *IGI global*, pp 242-264. 2010.

[14] Y. Pathak et al., "Deep transfer learning based classification model for COVID-19 disease," *Irbm*, Vol. 43, Issue 2, pp. 87-92, April 2022.

[15] Z.N.K Swati et al., "Brain tumor classification for MR images using transfer learning and fine-tuning," *Computerized Medical Imaging and Graphics*, vol. 75, pp 34-46, 2019.

[16] Z. Alyafeai, M.S. AlShaibani, and I. Ahmad, "A survey on transfer learning in natural language processing," *arXiv preprint arXiv:2007.04239,* 2020.

[17] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458,* 2015.

[18] Z. Li et al., "A survey of convolutional neural networks: analysis, applications, and prospects," *IEEE transactions on neural networks and learning systems*, vol. 33, no. 12, pp. 6999-7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827.

[19] W. Wang and J. Gang., "Application of convolutional neural network in natural language processing," *2018 International Conference on Information Systems and Computer Aided Education (ICISCAE)*, IEEE, 2018.

[20] A. Jiao, "An intelligent chatbot system based on entity extraction using RASA NLU and neural network," *Journal of physics: conference series*, *IOP Publishing*, vol. 1487, no. 1, 2020.

[21] K. C. Bandhu et al., “Health Care Chatbot using Natural Language Processing with SGD and ADAM Optimizer Parameter Optimization,” *In 2022 IEEE World Conference on Applied Intelligence and Computing (AIC)*, IEEE, pp. 136-142), 2022.

[22] M. Maryamah, A. Z. Arifin, R. Sarno, dan Y. Morimoto, “Query expansion based on Wikipedia word embedding and BabelNet method for searching Arabic documents,” *Int. J. Intell. Eng. Syst.*, vol. 12, no. 5, pp. 202–213, 2019.

[23] F. Sakketou dan N. Ampazis, “A constrained optimization algorithm for learning GloVe embeddings with semantic lexicons,” *Knowledge-Based Syst.*, vol. 195, pp. 105628, May 2020.

[24] M. Esposito, E. Damiano, A. Minutolo, G. De Pietro, dan H. Fujita, “Hybrid query expansion using lexical resources and word embeddings for sentence retrieval in question answering,” *Inf. Sci. (Ny).*, vol. 514, pp. 88–105, 2020.

[25] Y. Liu, K. Li, D. Yan, dan S. Gu, “A network-based CNN model to identify the hidden information in text data,” *Phys. A Stat. Mech. its Appl.*, hal. 126744, 2021.

[26] J. Sharma et al., "A novel long term solar photovoltaic power forecasting approach using LSTM with Nadam optimizer: A case study of India," *Energy Science & Engineering*, vol. 10, pp. 2909- 2929, 2022. doi:10.1002/ese3.1178

[27] KTB, PISS, and TIM Dakwah Pesantren. *Tanya Jawab Islam: Piss KTB*. Daarul Hijrah Technology, 2015.