Expanding the Breadth of Ability in Artificial Intelligence Systems with Decision Trees

Andrew McInnis Jr[1], Mohammad Alshibli[1], Ahmad Alzaghal[1] & Samir Hamada[2]

[1] Department of Computer Systems, State University of New York, Farmingdale State College, United States

[2] Department of Computer Science, Central Connecticut State University, United States

Correspondence: Mohammad Alshibli, Department of Computer Systems, State University of New York, Farmingdale State College, United States.

Abstract

This paper introduces a unique perspective. Rather than focusing on improving the already significant achievements of existing artificial intelligence algorithms, it investigates the potential of merging various algorithms to enhance their overall capabilities. Essential design aspects required for this integration are examined, and a prototype system is developed to demonstrate the practical application of these design principles. This method aims to broaden the range of capabilities accessible to a system, addressing the limitation of the narrow focus prevalent in contemporary artificial intelligence.

1. Introduction

Modern artificial intelligence is considered narrow (Schatz, 2018). An algorithm is entirely incapable outside the severely limited scope of what it has been trained to do, and it can only be trained to do one thing. That one thing is also limited in how broad it can be. Though general artificial intelligence may yet lie on the horizon, expanding the scope of what narrow artificial intelligence systems can do remains a contemporary issue. (Auger, Jacobs, Dobson, Marshall, & Noyce, 2020)

Algorithms are good at specific tasks, but they are not good at generalizing to new situations (Marcus, 2018). Machine learning systems are typically designed to perform a single task or a set of closely related tasks (Mitchell, 1997). The performance of a machine learning algorithm is heavily dependent on the quality and quantity of the training data it is given (Heaton, 2018a). Algorithms can only learn what they are exposed to in the training data (Domingos, 2015). Machine learning algorithms are often unable to extrapolate beyond the range of data they have been trained on (Pearl & Mackenzie, 2018). Algorithms are good at interpolation, but they are not good at extrapolation (Marcus & Davis, 2019). Machine learning algorithms can be biased and can make errors, even when they are performing well on the tasks they were trained on(Crawford, 2021). Algorithms are not perfect, and they can be fooled by adversarial examples(Goodfellow, Shlens, & Szegedy, 2014). The performance of a machine learning algorithm is heavily dependent on the quality and quantity of the training data it is given and the breadth of what can be considered one single task, there is relatively little in the way of expanding the scope of ability beyond singular tasks(Heaton, 2018b).

The approach in this paper is a system that incorporates multiple different artificial intelligence algorithms into a single cohesive unit. Each algorithm in the system is narrow, still capable of only one task, but activated at the discretion of another algorithm whose purpose is to decide how to solve the problem. This decision-making algorithm will identify a problem and select an appropriate algorithm to use for approaching it. In short, using an algorithm specialized in identifying a different algorithm most appropriate for the task at hand would significantly expand the scope of what an artificial intelligence system would be capable of. The rest of this paper is organized in the following way: Section 2 discusses previous work related to this paper, Section 3 discusses both the design method proposed (section 3.1) and the example system created (section 3.2), and Section 4 contains conclusions and considerations for future work.

2. Related Work

Decision trees such as those explored by (Charbuty & Abdulazeez, 2021) are subject to a great variety of research. (Gavrylenko, Sheverdin, & Kazarinov, 2020) uses various types of these algorithms to identify potential problems in computer systems, and (Suresh, Udendhran, & Balamurgan, 2020) uses a decision tree in combination with other tech- niques for complex medical diagnoses. (Y. Yuan & Shaw, 1995) discusses the fuzzy decision tree suggested by this paper, and (X. Yuan et al., 2019) combines one such algorithm with other techniques to increase stability and accuracy.

Voice cloning and synthesis are used by the example system in this paper. (Huang et al., 2020) explores an approach for rapid development.(Neekhara, Hussain, Dubnov, Koushanfar, & McAuley, 2021) suggests a method for creating more expressive voices. Impressive results in singing synthesis were achieved by (Blaauw, Bonada, & Daido, 2019; Shiga, Ni, Tachibana, & Okamoto, 2020). Explores text-to-speech techniques, which are necessary for any verbal communication requirements like those of the example system in section 3.2. Some techniques for more convincing voice cloning are suggested by (Sokolov, Alimov, Tyapkin, Katorin, & Moiseev, 2020).

The example system as a whole is similar conceptually in some ways to the agent created by (Poh et al., 2021). Other necessary components to the example system in this paper also have a wealth of research available. (Karri & Kumar, 2020; Skrebeca et al., 2021) explore techniques for creating chatbots. An AI approach to playing video games is proposed by (Kunanusont, Lucas, & Peŕez-Lieḃana, 2017). Generating commentary for such content with AI is explored by (Li, Gandhi, & Harrison, 2019). Closed captioning is explored by (Chen et al., 2019; Pavitha et al., 2023) and examines techniques for sentiment analysis.

The uncanny valley, and natural movement for animated conversational agents, have been explored thoroughly. (Geller, 2008) discusses the impact of style on this phenomenon, and (Cassell & Thorisson, 1999) found that body language was critically important for communication even with such agents. The humanness of AI agents was found to be important by both (Cheng, Zhang, Cohen, & Mou, 2022; Sheehan, Jin, & Gottlieb, 2020). (Ciechanowski, Przegalinska, Magnuski, & Gloor, 2019) touched on the issue of human response to chatbots and (Hill, Ford, & Farreras, 2015) produced interesting results on how these interactions play out. (Lu, Shen, Li, Shen, & Wigdor, 2021) explores similar topics relevant to the example system in this paper.

3. Discussion

3.1 Design Method

The component which makes up the system can be given 4 roles. Each component piece of the system should fill only one role and have only one purpose within the greater whole. Though components must rely on each other to function as a system, they should not be dependent on each other to function in their role. The reason for this is similar to that of replaceable parts. Single components of the system are easily updated in isolation or replaced with a new part that functions differently.

The first of the four roles components within the system may have is a sense. To operate in and respond to the environment in which it is situated, the system needs to be able to perceive it. Not all aspects of the environment are relevant to the system, such as an autopilot system not caring much about the highway traffic in the local area. The aspects the system needs to be able to sense should be well defined, however, and are specific to the purpose of a particular system. Each sense should be responsible for a certain type of data but should not be processing anything. That role should be left to another component.

The second of the four roles that components within the system may have is processing data from the senses. In addition to the raw data, there may be metadata about what sense picked up the data among other things. Each processing com- ponent should be dedicated to a single type of processing, such as sentiment analysis, natural language processing, image processing, or other types of data processing. In some systems, it may be necessary to have multiple stages of processing, such as generating closed captioning and then natural language processing on the captions, depending on the design goals in place. The final result of the data processing stage of the system's operation should be informed that the system can be used to make decisions. This information, or data with added context and understanding, should then be passed on to the next type of component along the line.

The third of the four roles that components within the system may have been to identify the problem to solve. The purpose of this role is to decide which Artificial Intelligence algorithm is the best fit for the problem at hand. It does this by classifying what the problem it is facing is, or if there is any problem at all. The information given to this component by the processing components must necessarily be organized into a data set that is usable by the particular Artificial Intelligence algorithm in use at this stage. There are several approaches to this issue that are possible.

The first approach is to segment data into short and long-term memory. Short-term memory is information that was recently gained, within a certain period of time defined by the designers of the system as necessary for the goals. Long- term memory is historical information, which was gained after the defined cut-off for short-term memory. The advantage to this is being able to consider experience in addition to immediate circumstances, however, in practice systems may not be able to accommodate the demands on processing time, memory and storage.

The second approach is to only use short-term memory. This means that any information beyond a certain age is lost to the system. The obvious disadvantage is data scarcity, which is a major concern for most Artificial Intelligence projects. The choice between the first and the second approach is largely dependent on the particular design goals of the system and what can be practically achieved by the designers. Inevitably, there is a finite amount of information that can be stored, no matter how large or small the database is.

With a usable data set for the Artificial Intelligence algorithm made for identifying the problem available, the question of what algorithm to use must be addressed. Deciding on which algorithm to use is fundamentally a classification problem, however, most things cannot be neatly categorized. Many problems encountered by a system in practice can be considered to be in multiple categories. For example, the problem of recognizing fast obstacles and avoiding them could be considered an issue of image processing, for recognizing specific obstacles, and as an issue of spatial awareness, for knowing which way to move to get out of the way. It is therefore necessary to consider degrees of membership to many different categories of problems.

In the July 2021 article (Nabil, Seyam, & Abou-Elfetouh, 2021) discusses the comparative advantages of decision trees over neural networks. Nabil in his paper highlighted the simplicity and ease of understanding offered by decision trees, as well as their superior ability to process categorical data, in contrast to the more complex neural networks.

While neural networks are certainly a strong choice, a decision tree is far easier to model and understand. Further, decision trees are better than neural networks in terms of handling categorical data. Since the data is fuzzy and not binary, a fuzzy decision tree (FDT) is the best choice of algorithm for this purpose. Techniques for increasing the stability and accuracy of classification by decision trees, such as bootstrap aggregation, may also be considered. The precise details of implementation depend largely on the specific requirements of the system being designed, though for most cases FDTs should be used.

The fourth and final role that a component can fill it in solving a problem, once identified. These artificial intelligence algorithms are specialized to solve specific problems, given the information passed to them by the other components of the system. It is therefore necessary to have a clear understanding of what problems the system is expected to encounter. In the case where a problem for which no algorithm is available to solve is encountered, the only reasonably expected result is a failure. Once this algorithm has determined a solution, action can be taken toward that end.
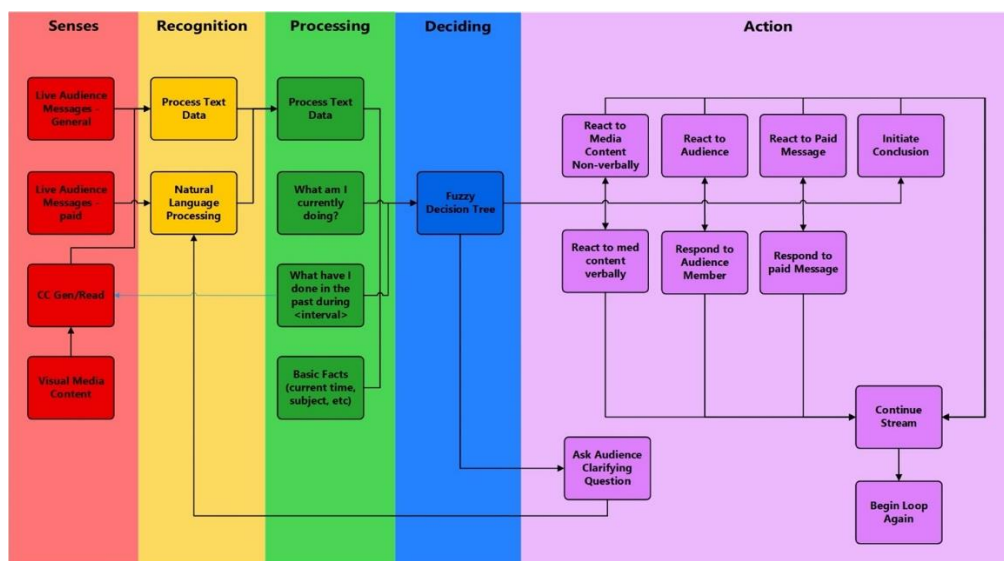
3.2 Example System

Figure 1. Proposed System

In order to demonstrate the outlined methodology for creating a multifaceted system of artificial intelligence algorithms, a conceptual design for one such system was created. This conceptual system needs to perform one of several different acts in front of a live audience using a digital medium called a stream. The different acts for this live stream include singing, playing video games or consuming other multimedia experiences, and providing commentary where appropriate.

The environment in which this system will reside has several components, in broad terms. The popular platform YouTube is used for delivering live streams to the audience. There is a live chat through which audience members can send messages that are displayed publicly, and these messages are delivered to the entertainer (streamer) as well. There are paid messages, highlighted in a color according to their cost, which are more visible than normal messages (super chats). In addition to the chat, the function is the subject of the stream, which typically takes up the most amount of screen real estate and is positioned to draw attention to it naturally. The streamers themselves are typically pictured from the chest up and front-facing in either the bottom-right or bottom-left corner of the screen.

The subject of the stream is typically chosen before the stream occurs and is not improvised. It is therefore possible, depending on the specific content, to prepare through preprocessing or some other means. Particularly, commentary on reliably predictable multimedia experiences such as movies or video content and singing are rarely improvised events. These particular problem-solving components of the system have less concern over timing since the product of their solutions can be saved for a scheduled time to execute them or play them back. Singing should not be interrupted midway through, however, commentary can be stopped at the end of thought even when the full commentary hasn't been delivered.

Video games as a subject are more difficult since they require more active engagement with the subject in a less predictable environment. While foreknowledge of events in the game may lend itself to commentary generation through preprocessing as above, the timing is not fully reliable in the vast majority of cases, so recognizing the moment to deliver any commentary needs to be through some other method. Recognizing a particular frame you know will occur is possible, but game- dependent and may not be reliable; Recognizing a particular voice line or other specific lines of closed captioning is another possibility, but in some cases may require closed captioning generation. Given that closed captioning generation is necessary in some cases with other stream subjects, and that natural language processing is also a necessary feature for processing some other data from the environment, the latter of the two options is the best choice.

The data being received from the environment through sensory components comes from a few sources. General audience messages and superchats, which should be treated differently and regarded as separate data sources, produce text data. The subject of the stream will produce visual and audio data, but in cases where timing and closed captioning are reliable as described earlier, both of these factors can be ignored during a live performance; For less predictable cases, such as video games, their importance is much higher. Closed captioning is not captured through a sensory component but rather is produced by a processing component.

The data from the environment goes to processing components. The metadata about the source and other relevant environmental factors can simply be passed along. Text data from general audience messages and superchats will go through natural language processing algorithms, similar in principle to a chatbot processing a message from a conversational partner. Sentiment analysis should be used to measure general audience reactions to things happening on the screen. Audio data will be used to generate closed captioning, before being similarly passed to a second processing step of natural language processing; natural language processing can be the same component for audience messages and superchats, but closed captioning is different fundamentally from human conversation, and so should be processed by a separate algorithm.

From processing the data, information is gained and can now be used to identify a problem. There is a case where there is no problem, or otherwise where no action is necessary. Further, concurrent problems may have solutions that cannot be done in parallel. Anything that requires verbalization, such as vocalizing a response to a message, providing commentary, or singing a song cannot be done simultaneously. Other tasks can be done in parallel, such as playing a game and responding to an audience message. The commentary must be paired with some subject, by its very nature. Identifying when errors have been made is also critical.

While no perfect failure recovery system can exist, a controlled crash approach is a good option in this case; Rather than correcting the mistake itself, correct course to compensate and minimize its impact. Recognition of failure or error, beyond a straightforward programmatic failure or error, can be recognized through stimuli in the environment. By detecting it through signs of its impact on the course of the live-streaming operation, failures and errors of negligible impact are naturally filtered out. Once recognized, what approach is taken is important

for the controlled crash approach to work.

In the case of a flubbed line, or one made out of an error in understanding, acknowledging the mistake verbally may be enough. Poor video game performance can simply become the object of commentary generation. Having low confidence in understanding stimuli can be met with non-articulate responses or asking for clarification where appropriate.

The problems that the system should expect to encounter are varied. Replying to messages with a synthesized voice, generating natural-sounding singing, filling the role of a player in a video game, reading the mood of the audience, adjusting things when there are negative reactions, and creating commentary are all issues that need to be handled. Each of these expected problems can in turn be solved by dedicated algorithms.

Some problems are more time-sensitive than others. In the case of a video game, an action sequence may demand attention over commentary. It is further important to define which issues are more important than others, in the case of simultaneous problems; A superchat takes precedence over a general audience message. However, there may be cases in which the system simply cannot address every problem. Dynamically choosing what to problems to address and what to ignore is an option, but in the case of a system such as this, a simple hierarchy of problems is an acceptable solution.

Further considerations for the system include aesthetic appeal and the attitude an audience might take toward an entity they know is a machine. More important than good artistic design is avoiding the uncanny valley. A stylized approach to design, rather than pursuing realism, is a sure method of success in this regard. Additionally, stylizing sound, or using another method like introducing background noise, to mask the distinctive intonation of computer-synthesized voices may be necessary. Since people have been found to have a proclivity for abusive behavior towards AI, it is advisable to either conceal the reality of this system or otherwise introduce a heavy element of computer-aided moderation to the live chat feature of the environment.

4. Conclusions and Future Work

While the final result is far from general intelligence, a system designed in the way described would have capabilities beyond that of a singular algorithm. The challenges inherent in any such project are difficult, but not insurmountable. State-of-the-art technology, skill, and vision are not enough to succeed. A clear map of how to proceed is necessary not only for the success of individual projects but also for communicating the processes effectively and for the success of other projects that follow.

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Open access

Copyrights

References

Auger, S. D., Jacobs, B. M., Dobson, R., Marshall, C. R., & Noyce, A. J. (2020). *Big data, machine learning, and artificial intelligence: A neurologist's guide.* Practical Neurology. https://doi.org/10.1136/practneurol-2020-002688

Blaauw, M., Bonada, J., & Daido, R. (2019). *Data efficient voice cloning for neural singing synthesis.* In 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 6840-6844). IEEE. https://doi.org/10.1109/ICASSP.2019.8682656

Cassell, J., & Thorisson, K. R. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence, 13*(4-5), 519-538. https://doi.org/10.1080/088395199117360

Chen, J., Pan, Y., Li, Y., Yao, T., Zhao, H., & Mei, T. (2019). Temporal deformable convolutional encoder-decoder networks for video captioning. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, pp. 8167-8174). https://doi.org/10.1609/aaai.v33i01.33018167

Cheng, X., Zhang, X., Cohen, J., & Mou, J. (2022). Human vs. AI: Understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms. *Information Processing & Management, 59*(3), Article 102940. https://doi.org/10.1016/j.ipm.2022.102940

Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental stu*dy* of human–chatbot interaction. *Future Generation Computer Systems, 92,* 539-548. https://doi.org/10.1016/j.future.2018.01.055

Crawford, K. (2021). The atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press. https://doi.org/10.12987/9780300252392

Domingos, P. (2015). The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World. Basic Books.

Gavrylenko, S., Sheverdin, I., & Kazarinov, M. (2020). The ensemble method development of classification of the computer system state based on decision trees. https://doi.org/10.20998/2522-9052.2020.3.01

Geller, T. (2008). Overcoming the uncanny valley. *IEEE Computer Graphics and Applications, 28*(4), 11-17. https://doi.org/10.1109/MCG.2008.79

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572.

Heaton, J. (2018a). Review of Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Genetic Programming and Evolvable Machines, 19*(1-2), 305-307. https://doi.org/10.1007/s10710-017-9314-z

Heaton, J. (2018b). Review of the book Deep learning, by I. Goodfellow, Y. Bengio, & A. Courville. *Genetic Programming and Evolvable Machines, 19*(1-2), 305-307. https://doi.org/10.1007/s10710-017-9314-z

Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real Conversations with Artificial Intelligence: A Comparison Between Human–Human Online Conversations and Human–Chatbot Conversations. *Computers in Human Behavior, 49,* 245-250. https://doi.org/10.1016/j.chb.2015.02.026

Huang, Y., He, L., Wei, W., Gale, W., Li, J., & Gong, Y. (2020). Using personalized speech synthesis and neural language generator for rapid speaker adaptation. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 7399-7403). https://doi.org/10.1109/ICASSP40776.2020.9053104

Karri, S. P. R., & Kumar, B. S. (2020). *Deep learning techniques for implementation of chatbots.* In 2020 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-5). [Location of Conference]. https://doi.org/10.1109/ICCCI48352.2020.9104143

Kunanusont, K., Lucas, S. M., & Pérez-Liébana, D. (2017). *General video game AI: Learning from screen capture.* In 2017 IEEE Congress on Evolutionary Computation (CEC) (pp. 2078-2085). https://doi.org/10.1109/CEC.2017.7969556

Li, C., Gandhi, S., & Harrison, B. (2019). End-to-end let's play commentary generation using multi-modal video representations. In Proceedings of the 14th International Conference on the Foundations of Digital Games (pp. 1-7). https://doi.org/10.1145/3337722.3341870

Lu, Z., Shen, C., Li, J., Shen, H., & Wigdor, D. (2021). More kawaii than a real-person live streamer: Understanding how the otaku community engages with and perceives virtual YouTubers. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1–14). https://doi.org/10.1145/3411764.3445660

Marcus, G. (2018). *Deep learning: A critical appraisal.* arXiv preprint arXiv:1801.00631. Retrieved from https://arxiv.org/abs/1801.00631

Marcus, G., & Davis, E. (2019). Rebooting AI: Building artificial intelligence, we can trust.

Mitchell, T. M. (1997). *Machine Learning.* McGraw-Hill.

Nabil, A., Seyam, M., & Abou-Elfetouh, A. (2021). Prediction of students' academic performance based on courses' grades using deep neural networks. *IEEE Access, 9,* 140731-140746. https://doi.org/10.1109/ACCESS.2021.3119596

Neekhara, P., Hussain, S., Dubnov, S., Koushanfar, F., & McAuley, J. (2021). Expressive neural voice cloning. In Asian Conference on Machine Learning (pp. 252-267).

Pavitha, N., Ratnaparkhi, P., Uzair, A., More, A., Raj, S., & Yadav, P. (2023). Explainable AI for sentiment analysis. In Ict with intelligent applications (429-439). Springer. https://doi.org/10.1007/978-981-19-3571-8_41

Pearl, J., & Mackenzie, D. (2018). The book of why: The new science of cause and effect. Basic Books.

Poh, S. C., Tan, Y. F., Ooi, C. P., Tan, W. H., Quek, A., Gan, C. Y., ... Cham, C. L. (2021). Alice: A general-purpose virtual assistant framework. In Computational Science and Technology (pp. 383-393). Springer. https://doi.org/10.1007/978-981-33-4069-5_31

Schatz, M. E. (2018). Enabling Composite Optimization Through Soft Computing of Manufacturing Restrictions and Costs via a Narrow Artificial Intelligence. *Journal of Composites Science, 2*(4), 70. https://doi.org/10.3390/jcs2040070

Sheehan, B., Jin, H. S., & Gottlieb, U. (2020). Customer service chatbots: Anthropomorphism and adoption. *Journal of Business Research, 115,* 14-24. https://doi.org/10.1016/j.jbusres.2020.04.030

Shiga, Y., Ni, J., Tachibana, K., & Okamoto, T. (2020). *Text-to-speech synthesis.* In Speech-to-speech translation (pp. 39-52). Springer. https://doi.org/10.1007/978-981-15-0595-9_3

Skrebeca, J., Kalniete, P., Goldbergs, J., Pitkevica, L., Tihomirova, D., & Romanovs, A. (2021). Modern development trends of chatbots using artificial intelligence (ai). In 2021 62nd international scientific conference on information technology and management science of Riga Technical University (itms) (pp. 1-6). https://doi.org/10.1109/ITMS52826.2021.9615258

Sokolov, S. S., Alimov, O. M., Tyapkin, D. A., Katorin, Y. F., & Moiseev, A. I. (2020). Modern social engineering voice cloning technologies. In 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EICONRUS) (pp. 513-516).

https://doi.org/10.1109/EIConRus49466.2020.9038954

Suresh, A., Udendhran, R., & Balamurgan, M. (2020). Hybridized neural network and decision tree-based classifier for prognostic decision making in breast cancers. *Soft Computing, 24*(11), 7947-7953. https://doi.org/10.1007/s00500-019-04066-4

Yuan, X., Wang, X., Han, J., Liu, J., Chen, H., Zhang, K., & Ye, Q. (2019). A high accuracy integrated bagging-fuzzy- gbdt prediction algorithm for heart disease diagnosis. In 2019 IEEE/CIC International Conference on Communications in China (ICCC) (pp. 467-471). https://doi.org/10.1109/ICCChina.2019.8855897

Yuan, Y., & Shaw, M. J. (1995). Induction of fuzzy decision trees. *Fuzzy Sets and Systems, 69*(2), 125-139. https://doi.org/10.1016/0165-0114(94)00229-Z