Introduction

This research paper will be looking at the incorporation of distributed big data architecture and deep learning neural networks to solve for specific use cases. The problem area being examined as part of the practical work is classifying urban sounds for the purposes of reducing noise pollution. By not only monitoring noise levels, but being able to identify the sources of the noise pollution allows for the creation of plans and policies designed to combat a specific issue that has been uncovered, thereby improving quality of life for the local community. Neural networks are well-suited to audio classification for a number of reasons including robustness and flexibility, and so were chosen as the machine learning model to be used in the analysis. Neural networks are also scalable, which is important for urban sound classification, and to achieve a similar scalability for data storage, a distributed big data architecture was necessary. Apache Kafka is used throughout the analysis for this purpose.

In addition to the practical deployment there will be a literature review performed on key papers which examine the topic of neural networks and big data architecture as it relates to urban sound classification and storage. As a result of the practical research and the literature review, a critical evaluation of the key findings will be presented, implications for future research and current limitations will be discussed, as well as knowledge gaps which could be addressed with further research.

1.1 Topic Overview

“Urban Sounds are everyday audio events that occur daily, presenting unstructured characteristics containing different genres of noise and sounds unrelated to the sound event under study, making it a challenging problem” (Nogueira et al., 2022)

There are a number of reasons why people attempt to solve the challenging problem of urban audio recognition, ranging from the chosen topic area of noise pollution, to public safety for gunshot or vehicle crash alerts. City optimisation with the detection of traffic congestion allowing for the adjustment of traffic lights or rerouting of lanes. All of these use cases depend on a reliable and accurate model which can predict the audio data. It can also require a large amount of computational power and storage. For those two reasons a common methodology is to use a neural network model, a distributed big data architecture, and often both in combination.

Neural networks “name and structure are inspired by the human brain” (IBM, 2021), logical when considering they will be used for audio recognition, which according to Hennebert (1994) is a complex perceptive task which other machine learning algorithms struggle with. Once created the model will intake audio and classify it based on previous training data. These predictions are what would help solve or answer the use cases mentioned above.

Storage of data is also an issue and audio data is much larger than numerical or text based data, meaning that a big data storage solution will be required. Distributed storage systems allow for storing large amounts of data as compared to traditional databases by being “designed to scale horizontally across clusters of low-cost, moderate-performance servers” (Gorton and Klein, 2015)

1.2 Objective Statement

Using a large set of labelled audio data relating to urban sounds, create and train a neural network model which can be used to classify future audio. Use the trained model to classify audio, and by means of a distributed big data architecture, store the audio recording, and the prediction result, for later use.

1.3 Research Question

How have neural networks been used to classify urban sounds and are there predominant methods or models used? Has big data architecture been used in the storage of urban sound data and if so which types of architectures were implemented. Is there missing research which could help address the problem area?

2. Materials

The models created in this analysis were trained on an urban sounds audio database, containing over 8000 labelled sound excerpts from 10 different classes of sounds. This was downloaded via a zip file from the original website, “<https://urbansounddataset.weebly.com/urbansound8k.html>” and then unzipped. The unzipped folder contains a csv file with metadata including the class labels for the audio files. There are then 10 separate folders each with over 80 audio files. Using Python, a function is created to extract the mel frequency cepstral coefficients (MFCCs), which are “highly effective in audio recognition and in modelling the subjective pitch and frequency content of audio signals” (Xu et al., 2004). The input is a file path and metadata, which will be the urban sound recordings, while the output will be the MFCCs and the class label. The function is then used in a for loop to process all of the audio clips into a ‘feature\_list’, containing the MFCCs and a ‘label\_list’ holding the corresponding class label. Both are multi-dimensional numpy arrays. These are then saved to disk to avoid having to reprocess them all on subsequent analysis’.

3. State of the Art

3.1 Research Methodologies

The research methodologies used in this paper included research design, data collection and data analysis.

The research design took a combined approach of quantitative and qualitative methods to answer the research questions. This allows for the performance of a verifiable analysis, which is then enhanced with findings from multiple sources and opinions generated from state of the art research. This is all with the aim of providing valid results that give insight on the topic being discussed and the research questions posed.

Data collection methods included an audio dataset previously compiled by other researchers for the purpose of carrying out studies such as this. It is 8732 audio clips of five seconds or less duration, all related to one of ten classes. The class labels are stored in a csv file which also contains other metadata around the audio files. There was then a document analysis performed in the form of a literature review. This consists of reviewing and analysing relevant documents such as reports, journal articles, conference papers or book excerpts. This is done to enhance the practical aspect and give either more support to its findings or reasons for further analysis.

The data analysis was performed on the audio dataset and took the form of exploratory data analysis (EDA), where descriptive statistics are generated to give insight to the data. Model creation and integration of Apache Kafka was performed to try to solve for a problem area and the results were all recorded.

3.2 Key Papers Reviewed

* Aletta, F., Kang, J., & Axelsson, Ö. (2016). Urban soundscape categorization based on individual recognition, perception, and assessment of sound environments.
* Giannakopoulos, T., Spyrou, E., & Perantonis, S. J. (2019). Recognition of Urban Sound Events Using Deep Context-Aware Feature Extractors and Handcrafted Features.
* Piczak, K. J. (2015). Environmental sound classification with convolutional neural networks.
* Bhole, R. H. (2018). A Study of Apache Kafka in Big Data Stream Processing.
* Bhole, R. H., & Hiraman, B. R. (2018). KAFKA: The modern platform for data management and analysis in big data domain.
* Bhole, R. H., & Hiraman, B. R. (2018). ESTemd: A Distributed Processing Framework for Environmental Monitoring based on Apache Kafka Streaming Engine.
* Salamon, J., Jacoby, C., & Bello, J. P. (2014). A dataset and taxonomy for urban sound research.
* Schedl, M., Flexer, A., & Urbano, J. (2013). Sound Classification and Processing of Urban Environments: A Systematic Literature Review.
* Salamon, J., & Bello, J. P. (2017). Urban Sound Classification: striving towards a fair comparison.

4. Methods

4.1 EDA

As part of the exploratory data analysis (EDA) basis information is gathered from the feature list, as well as the data frame created to represent the MFCCs and classes of each audio clip.After displaying the distribution of the class labels a t-test was performed to see which MFCCs had a statistically significant difference between the children playing and dog barking classes.

More visuals and statistical analysis is performed describing the dataset, and then anova tests are performed to list classes which do have statistically significant differences.

4.2 Sound Classification Model

A neural network model was decided for the project and it was implemented using the tensorflow and keras packages. For all the models the feature and label lists are converted to numpy arrays and the class labels are encoded. The labels are encoded as keras models cannot directly process categorical labels. The data is then split into training and testing sets with an 80/20 split.

The first model created used an input layer taking the feature vectors as inputs. The next layer is a dense layer with 128 neurons and the ‘relu’ activation function. The final layer is the output layer which has 10 neurons, one for each class and the ‘softmax’ activation functions. The model is compiled using the ‘adam’ optimiser, and then trained using the ‘fit.()’ method on the training data.

The next model uses a different structure but is still made up of input, output and hidden layers. In this case there are two dense layers, both using the ‘relu’ activation, with 256 and 128 neurons respectively. The model is compiled and fit in the same way as the first.

The final model was created using 10-fold cross-validation which splits the data into 10 folds and then trains on 9 and tests on 1, repeating the process for each fold. The model used is the same as the second with 2 dense layers. Each of the 10 models are compiled and fit in the same way as the first two models.

For all models classification reports are printed and Training and Validation accuracy line charts are visualised.

4.3 Kafka Topic Storage

After installing Apache Kafka on the machine and starting the Zookeeper and Kafka servers, Python code is run which records a predefined length of audio which is split into smaller chunks. These are then processed to extract the MPCC features which are then run through the saved neural network model in order to make a prediction. The audio clip and prediction are then sent to a Kafka topic as key pair values using the ‘KafkaProducer()’ function. Once the specified duration has been reached the process will terminate itself.

To verify that the process has worked, a prediction has been made and both it and the recording were successfully saved to the topic, another function, ‘KafkaConsumer()’ is used. This allows for the retrieval of messages from a topic. As the format of the stored data is known, the code includes functions for decoding the utf-8 encoded prediction string, as well as reading the audio back to a playable format from the bytes it was stored as.

5 . Literature Review

5.1 Urban Sound Classification

A sometimes challenging but important task, there is a good amount of up to date literature on urban sound classification as it has application in many areas as has been mentioned in the topic overview. There are barriers to the exploration of the area however and according to Salamon, Jacoby and Bello (2014), one such barrier is the lack of a common taxonomy, that is a stand scheme of classification, and beyond that there isn’t enough real-world annotated date to perform the classification tasks. Solving this involved creating the UrbanSound database, “containing 27 hours of audio with 18.5 hours of annotated sound event occurrences across 10 sound classes” (Salamon, Jacoby and Bello, 2014). This is the dataset used for the analysis in this paper.

This dataset has since been used in a number of papers including Arnault, Baptiste Hanssens and Riches (2020) ‘Urban Sound Classification: striving towards a fair comparison’ where they looked at creating a model for noise pollution. As the field of urban sound classification is growing however, their paper also had a secondary purpose, providing a solution for reproducing and comparing these type of models, with the aim that the “framework could help evaluate new architectures in this field” (Arnault, Baptiste Hanssens and Riche, 2020). As it is a developing field with new changes and research being carried out at a rapid pace, there are attempts to summarise the most common factors that would need to be considered when dealing with the topic.

‘Sound Classification and processing of Urban Environments: A Systematic Literature Review’ is one such paper and looks at the most recent works in order to “understand the current approaches and identify their limitations'' (Massoudi, Verma and Jain, 2021). This paper was mainly focused on how to create an efficient sound model, and so was concerned with the techniques that returned the best results. Ultimately data augmentation techniques and pretraining were found to be among the most important factors to consider.

5.2 Neural Networks

When considering neural networks, as Massoudi, Verma and Jain (2021), referred to, there are a number of factors that need to be considered. While the cutting edge is constantly experimenting and testing different techniques, at this point there is a standard audio classification process using deep learning which can be used as a template for then tuning and experimentation for increased results. This is the model we will be following for the analysis and it involves using a convolutional neural network, or CNN, and is trained of Mel Frequency Cepstral Coefficients (MFCCs) as described by Massoudi, Verma and Jain (2021) in their work on ‘Urban Sound Classification using CNN’.

While this is a straightforward method of classification, in the pursuit of better models, many other factors can be taken into account, including the context of the audio source. Passing this knowledge into “deep context-aware feature extractors” (Θεόδωρος Γιαννακόπουλος and Stavros Perantonis, 2018), is the purpose of a paper written around this, and the approach is taken to use CNNs as a method to “extract context-aware deep audio features that can offer supplementary feature representations to any soundscape analysis classification task”. Combining “scene” samples, such as park, or train, with handcrafted audio features resulted in a significant performance boosting. Layering more data or augmenting the existing data is a large opportunity for future research.

‘Urban Sound Tagging using Convolutional Neural Networks’ looks at how data augmentation techniques such as “Mixup, Random erasing, scaling, and shifting” (Sainath Adapa, 2019) are used to deliver “higher performance over alternative approaches”. This was used in a low data context with less than 100 labelled examples per class.

5.3 Big Data Solutions

Big data is a topic that has been researched for many years and there is a large volume of information around it. As the analysis in this paper uses Apache Kafka, the papers reviewed will also focus on this particular tool, although the domain of big data, even as it relates to urban sound classification, is much wider.

Kafta is a very popular architecture and so there is a large community around it and documentation available. In ‘A Study of Apache Kafka in Big Data Stream Processing’, Hiraman, Viresh M. and Abhijeet C. (2018) look at why in the current era Apache Kafka is the first choice as an architecture. They find that it delivers on the key needs of the user for “scalable, distributed and reliable results” (Hiraman, Viresh M. and Abhijeet C., 2018). The paper provides a useful introduction and overview of Kafka.

When it comes to the use cases of Kafka and even how it operates, using producers and consumers, with brokers and partitions, Shree et al. (2017) cover how these are implemented, with the paper again emphasising the scalability of Kafta where “a single cluster [can] serve the central data backbone for a large organisation” (Shree et al., 2017).

Because of this distribution and scalability, Kafka can be used for many different projects. In ‘ESTemed: A Distributed Processing Framework for Environmental Monitoring based on Apache Kafka Streaming Engine '' Akanbi (2020), Kafka is used as the basis for a proposed distributed framework, designed specifically around heterogeneous environmental data. This in turn would be used for “environmental decision support systems, early warning and forecasting systems'' (Akanbi, 2020). Using Kafka as part of sound recognition is a key part of the study done as part of this paper, and it appears to be a logical fit for sound recognition based on the focus of the paper.

6 .Critical Analysis

6.1 Audio Classification with Neural Networks

6.1.a Implications

The implication of Neural Networks accurately and reliably predicting urban sounds can be wide ranging. In the papers look at the was an initial model created to demonstrate the theory behind the process, and how with tools like Python and Apache Kafka it can be relatively straightforward. The implications become greater when the use cases become more specific, or the methods more complex. Θεόδωρος Γιαννακόπουλος and Stavros Perantonis (2018) attempted to involve context features, in this case the location of the urban sound event, for example park, or alley. This was then combined with hand crafted audio features to create a model which then classified an unknown audio recording. The purpose being to remove the need for CNN training a “demanding process that requires huge datasets and complex data augmentation procedures” (Θεόδωρος Γιαννακόπουλος and Stavros Perantonis, 2018).

Another paper takes an even more abstract approach in that it classifies observers of sounds events and interprets how they respond to certain audio events depending on their location in a room Jo and Jeon (2021). This kind of soundscape analysis is becoming more popular and another area where neural networks are used for audio classification.

6.1.b Limitations

Limitations around neural networks and the implementation of urban sound recognition lie in two areas identifiable by looking at the research papers. The first is data availability. Looking at three papers which use audio datasets as part of their analyses, two use the same dataset, UrbanSounds8k, the same used in this analysis, while the other uses a custom data set of fewer than 100 records. This lack of large, diverse, labelled datasets restricts the modelling that is possible and increases the likelihood of failure when put into production. Community efforts such as the UrbanSounds8k dataset have attempted to address the issue, but it is still a relatively niche category. As there are so many use cases that would provide substantiation benefit were they to succeed, the argument for creating more datasets is becoming more persuasive, and as the research and knowledge grows so will the amount of data available. It is worth noting that this organic growth may slow the development of the area.

6.1.c Contradictory Viewpoints

Contradictory viewpoints exist to the extent that in different papers, different neural network models were used with different parameters. However this is not contradictory as each paper was looking at a specific use case and so they are not comparable in that sense. Neural networks in the current literature seem to be the default for audio recognition, in the same way they are for image recognition.

6.1.d Research Gaps

While every attempt has been made to cover the state of the art, some articles may have been missed. Therefore any research gaps identified in this paper are based on current knowledge and are not intended to be comprehensive. The lack of data has been covered as a limitation, although it could also be improved by further research.

In the paper on ‘Urban soundscape categorization’ Jo and Jeon (2021), the issue is raised that more diverse and realistic urban environments and sound sources should be used. They also argued for organising the study “over a long period through repeated-longitudinal design” (Jo and Jeon, 2021). Although they are writing this specifically for their paper it is also true of the urban sound recognition area overall in that while there is some research and proposals out there, more long term research is needed to prove some of the ideas being tested.

6.2 Big Data Architecture Solutions.

6.2.a Implications

All of the articles highlight the importance that Kafka plays in the overall ecosystem of big data and the ability it has in terms of handling large volumes and diverse formats of data. For the use case of urban sound classification it enables the storage of audio files, and text files, using the same producer to push them to a topic. Sound classification is often an ongoing process in that it needs to run constantly in order to achieve its objectives. The benefits the papers highlight around Kafta match the big data processing needs of urban sound classification. In the case of the study for this paper, the “key, value and timestamp” (Shree et al., 2017) format of the Kafka messages, suited the predicted value, audio file pairing that needed to be stored. The suitability of Kafka for this task has led to the proposal of further developing a distributed processing framework, specifically for environmental monitoring and based around Apache Kafka. This would attempt to “offer real-time processing of huge environmental datasets” (Akanbi, 2020). This is one way that Kafka could positively affect the overall area of urban sound recognition, but it and more still have potential to develop systems that could solve numerous use cases.

6.2.b Limitations

The limitations of using Kafka are mainly around the technical difficulty of setup and pipeline building. If it is the correct architecture to solve for the particular use case, there don’t appear to be technical limitations that were run up against as part of this paper. Kafka provided the ability to achieve the research objectives, the success of the study depended more on the execution of the ideas and technical ability. While Kafka does require advanced technical knowledge to set up and maintain, there is a variety of documentation around it.

6.2.c Contradictory Viewpoints

In the research articles referenced in this paper, there was not a large amount of disagreement of the optimal usage or purposes of Apache Kafka. It’s recognised as a flexible big data architecture and in the practical analysis performed the implementation and results aligned with what the research reported Kafka to be. There are a number of different ideas about how it can be used as an architecture, both within urban sound recognition at a larger scale, but this is complimentary work not contradictory.

6.2.d Research Gaps

Two areas of research that would benefit the research around urban sound recognition would be data security and governance as related to Kafka, and a comparison between other major big data architectures, based on metrics like performance, scalability, reliability and usability.

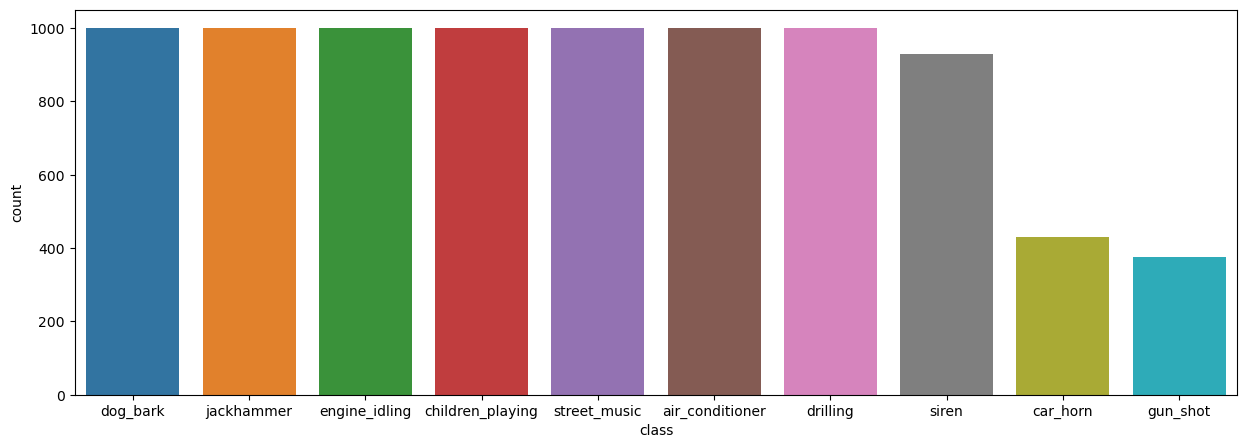
If large amounts of sound data taken from populated urban areas are being processed by Kafka, there would naturally be a concern around the safety of that data, and the effect any exposure may have. How Kafka handles data security is something that is documented but not researched in this specific area.

Having a comparison of Kafka with other big data architectures, specifically related to urban sound recognition, would give the opportunity to understand if Kafta is the gold standard, or if there are areas that other architectures do better, and what we could learn from that, whether it's migrating to a different architecture, or attempting to replicate that which it does better.

7. Results

7.1 EDA

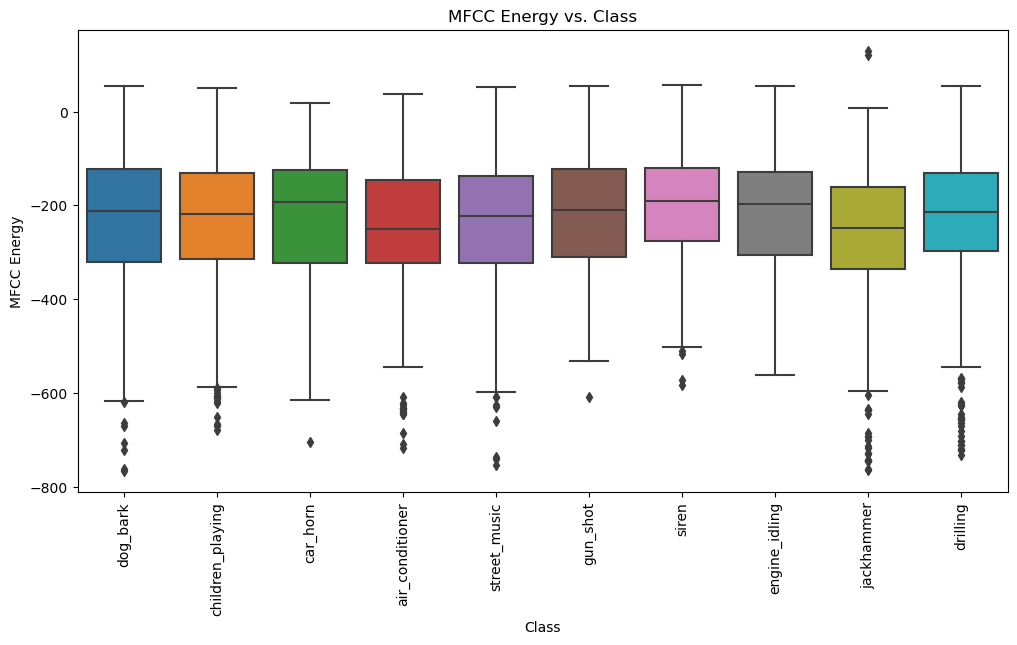
There are 8732 audio samples each with 13 different MFCC features. Examining the distribution of class labels shows 7 of the class labels have 1000 samples while ‘car\_horn’ and ‘gun\_shot’ have considerably less. This could affect the modelling so it will have to be taken into account.



The t-test showed that all features except for ‘mfcc\_spectral\_low’ had statistically significant differences in their means. These differences allow for the classification of different sound types.

The results of the ANOVA testing shows that all classes have differences in the means of at least one of its feature vectors. For most classes more than half of their features have statistically significant differences.

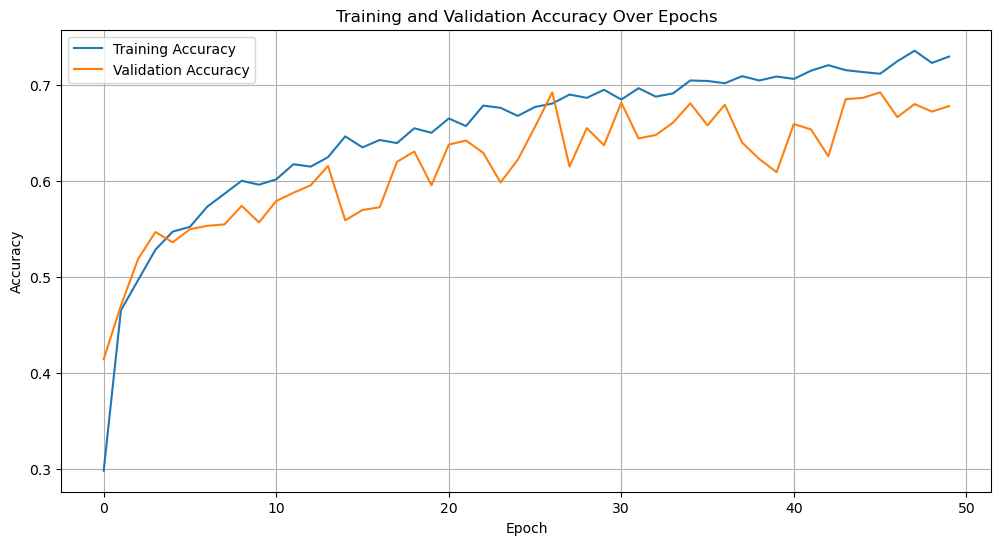
These variations can be seen when comparing a specific feature across classes, such as the ‘mfcc\_energy’ feature. Boxplots show that while some appear to have similar means, many classes vary significantly from one another.



7.2 Neural Network Model

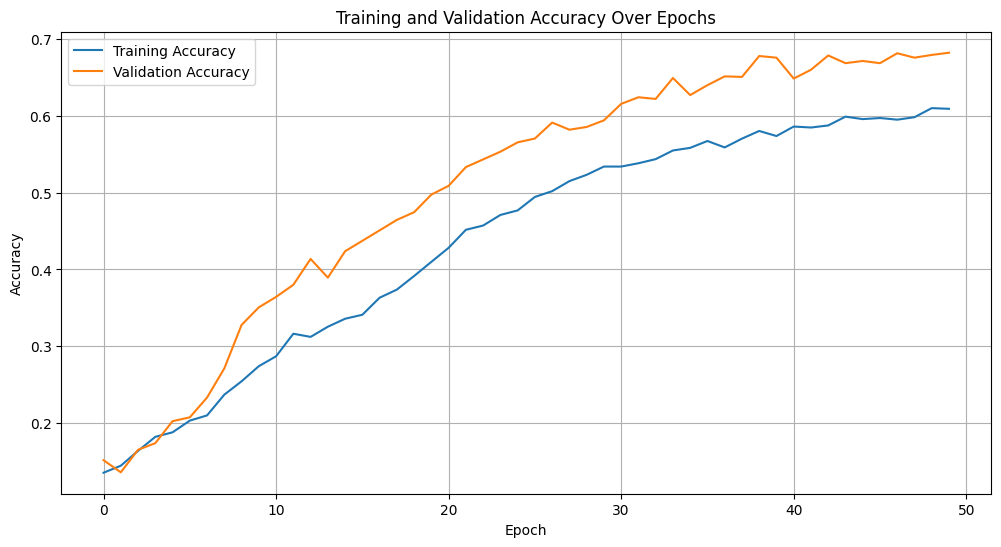
The first model returned a training loss of 0.76 with an accuracy of 0.74 on its final epoch.

The difference between the training and validation accuracy over time shows that the training accuracy is not significantly higher than the validation accuracy, which would be an indication of overfitting.

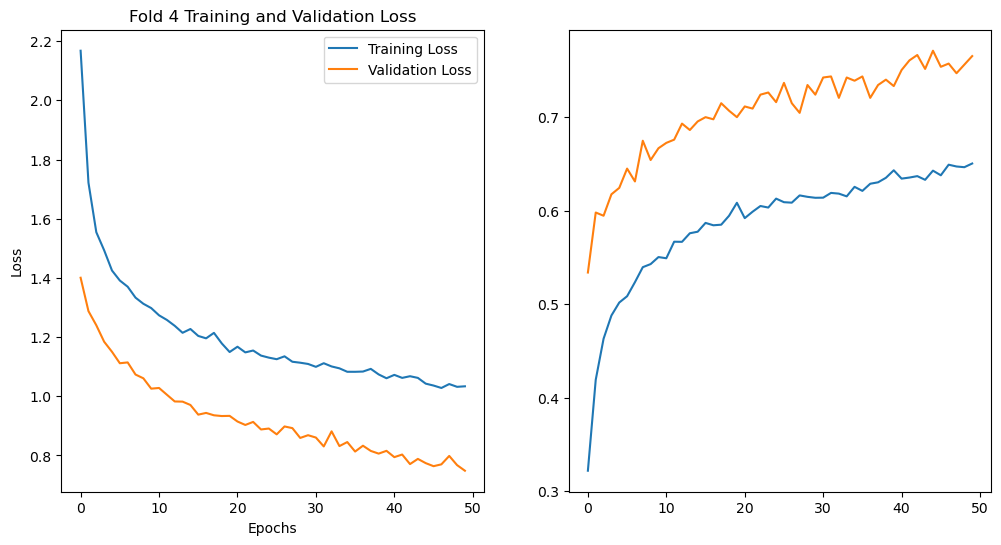


The classification report shows that ‘jackhammer’ and ‘siren’ are the classes with the highest f1-score, which balances precision and recall. These would be two loud and consistent noises so it is logical that they classify well as opposed to ‘street\_music’ for example which can be so varied that it is difficult to account for, resulting in low precision and recall results.

The second model results in training loss of 1.13 and an accuracy of 0.61 after its final epoch. This is a poorer performance than the first model. When comparing the validation results, validation loss is lower and accuracy is higher than the training results. This could indicate underfitting, however the differences are not extreme.



The third approach using 10-fold cross validation results in 10 separate models. Fold 4 has the best results with a training loss of 0.69 and a training accuracy of 0.78.



7.3 Kafka

The recordings and predictions were successfully sent to the topic, as verified by using kafka-consumer-console to view the messages. In terms of the results, during practical run throughs, when presented with sounds matching one of the class labels, the model has been able to classify some sounds and store this prediction as text to the kafka topic. The recording is also successfully stored, however on retrieval the audio is somewhat distorted. Listening to the audio before sending shows that it is clear and without distortion, so this occurs either when encoding the audio to be sent or decoding it on retrieval. As the model processing happens before encoding/sending the predictions are being made on clear audio, but this is an area where further work could significantly improve the quality.

8.1 Key Findings

Urban sound recognition is a growing field with a number of interesting contributions to date. When attempting to classify audio data, neural networks perform well in classifying extracted features, in this case MFCCs. In the practical testing a model with 2 hidden layers, using the ‘adam’ optimiser gave the best results. Apache Kafka is a powerful big data architecture that can be used to stream data to topics in key pair values, and was particularly suited for dealing with the mix of audio and text data that was being produced as part of the analysis. It is also scalable and therefore capable of being put into widespread use.

The results of the model were not as accurate as would have been desired however, and there was some issue with audio quality after retrieval using Kafka.

8.2 Implications

The practical analysis shows a proof of concept for recording large amounts of audio data in usable pieces, classifying them using a trained CNN, and then storing them, all in real-time. This could be a powerful tool for sound recognition in the case of gunshots or car crashes, as well as other examples we’ve previously given. While it still requires a good deal of development, if it were to be put to use and solve or reduce the problem area, it could be of great benefit.

8.3 Future Research

Future research for this paper would involve looking at more ways to improve neural network accuracy, including data augmentation techniques. There were a number of suggestions on how to improve model performance and some of these could be used or tested as part of the practical analysis.

There would also need to be more research done into how to use Kafka in the most optimal way. If it is going to be processing high volumes of data, ensuring it is setup to run as efficiently as possible is extremely important. More investigation into distributed processing frameworks would also be beneficial when looking to scale the application.

9. Conclusions

Using a large dataset of urban sounds, UrbanSounds8K, features were extracted to perform classification modelling. EDA was performed on the features to understand the nature of the different audio classes being examined, and visuals were used to assist in interpreting the outputs. Using the keras and tensorflow packages, neural network models were trained on the labelled data in a variety of ways including sequential CNNs and using 10-fold cross-validation. The models were then tested against a test set of data and the results were printed in the forms of confusion matrices, classification reports, and accuracy and loss visualisations. The models were also saved. Audio data was then streamed, classified and stored with its classification as a key pair value using Apache Kafka. These were then retrieved as part of the same example. The findings showed that a 2 layer neural network gave the best performance.

As part of the part a literature review was also conducted to support and expand on the practical analysis. Nine specific papers were chosen to represent the areas of the practical investigation and they were reviewed and analysed to help inform the findings.

Neural networks and Apache Kafka were both confirmed as good choices for the work by the literature review, however there was significant scope to improve the performance of both in the practical analysis, with the assistance of further research.

Overall the area is one that is growing and has a high potential to positively impact a number of initiatives.

Reference List

Akanbi, A.K. (2020). ESTemd: A Distributed Processing Framework for Environmental Monitoring based on Apache Kafka Streaming Engine. *arXiv (Cornell University)*. doi:https://doi.org/10.1145/3445945.3445949.

Arnault, A., Baptiste Hanssens and Riche, N. (2020). Urban Sound Classification : striving towards a fair comparison. *arXiv (Cornell University)*.

Gorton, I. and Klein, J. (2015). Distribution, Data, Deployment: Software Architecture Convergence in Big Data Systems. *IEEE Software*, 32(3), pp.78–85. doi:https://doi.org/10.1109/ms.2014.51.

Hennebert, J. (1994). Neural Networks In Speech Recognition. [online] 6th Microcomputer School. Available at: https://www.researchgate.net/publication/2249623\_Neural\_Networks\_In\_Speech\_Recognition [Accessed 26 Sep. 2023].

Hiraman, B.R., Viresh M., C. and Abhijeet C., K. (2018). *A Study of Apache Kafka in Big Data Stream Processing*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICICET.2018.8533771.

IBM (2021). *What are Neural Networks? | IBM*. [online] www.ibm.com. Available at: https://www.ibm.com/topics/neural-networks.

Jo, H.I. and Jeon, J.Y. (2021). Urban soundscape categorization based on individual recognition, perception, and assessment of sound environments. *Landscape and Urban Planning*, 216, p.104241. doi:https://doi.org/10.1016/j.landurbplan.2021.104241.

Massoudi, M., Verma, S. and Jain, R. (2021). *Urban Sound Classification using CNN*. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICICT50816.2021.9358621.

Nogueira, A.F.R., Oliveira, H.S., Machado, J.J.M. and Tavares, J.M.R.S. (2022a). Sound Classification and Processing of Urban Environments: A Systematic Literature Review. *Sensors*, 22(22), p.8608. doi:https://doi.org/10.3390/s22228608.

Nogueira, A.F.R., Oliveira, H.S., Machado, J.J.M. and Tavares, J.M.R.S. (2022b). Sound Classification and Processing of Urban Environments: A Systematic Literature Review. *Sensors*, 22(22), p.8608. doi:https://doi.org/10.3390/s22228608.

Sainath Adapa (2019). Urban Sound Tagging using Convolutional Neural Networks. *arXiv (Cornell University)*. doi:https://doi.org/10.48550/arxiv.1909.12699.

Salamon, J., Jacoby, C. and Bello, J.P. (2014). A Dataset and Taxonomy for Urban Sound Research. *Proceedings of the ACM International Conference on Multimedia - MM ’14*. [online] doi:https://doi.org/10.1145/2647868.2655045.

Shree, R., Choudhury, T., Gupta, S.C. and Kumar, P. (2017). KAFKA: The modern platform for data management and analysis in big data domain. *2017 2nd International Conference on Telecommunication and Networks (TEL-NET)*. doi:https://doi.org/10.1109/tel-net.2017.8343593.

Θεόδωρος Γιαννακόπουλος and Stavros Perantonis (2018). Recognition of Urban Sound Events using Deep Context-Aware Feature Extractors and Handcrafted Features. *HAL (Le Centre pour la Communication Scientifique Directe)*. doi:https://doi.org/10.20944/preprints201811.0509.v1.

Xu, M., Duan, L.-Y., Chia, L.-T. and Xu, C. (2004). Audio keyword generation for sports video analysis. doi:https://doi.org/10.1145/1027527.1027702.