# Learning Audio Features for Music Classification and Recommendation with Deep Learning

*Tinlok Pang (Dr Fabio Ramos)*
School of Information Technologies
FACULTY OF ENGINEERING & INFORMATION TECHNOLOGIES

## Motivation and Problem

### Music in the 21st Century

- There has been an explosion in the ease of accessibility and distribution of music over the last 15 years, due to the Internet facilitating peer-to-peer file sharing services and online music stores.
- These have driven the popularity of *music recommendation systems,* as more of the world's population seek to discover new music they may not have had access to before.

### Music Recommendation Systems

- Music recommendation systems take as input a set of songs that a user enjoys, and attempts to return a *recommendation* of a completely new set of songs that the system *predicts* the user may also enjoy and rate highly.
- These *recommendations* are essentially songs and music that have been *classified* in some way to be likely to agree with the user.
- Thus, *music recommendation* is a *classification* problem.

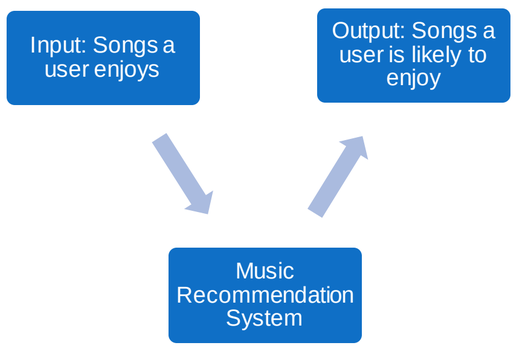



*Fig 1: Music recommendation systems assist users in finding new music they might enjoy*

### Classification

- Most classification systems require the use of human perception to either observe audio features to be computationally measured or as the primary measure of such features.
- This is problematic as it limits the music classification features to only concepts that humans can imagine and implement.

## Method

### Automatic Feature Learning and Deep Learning Architectures

- Deep learning architectures are a complex network of layered non linear operations as shown in *Fig 3*.
- These architectures can be trained on a dataset to learn it's salient features automatically, with no human interference.

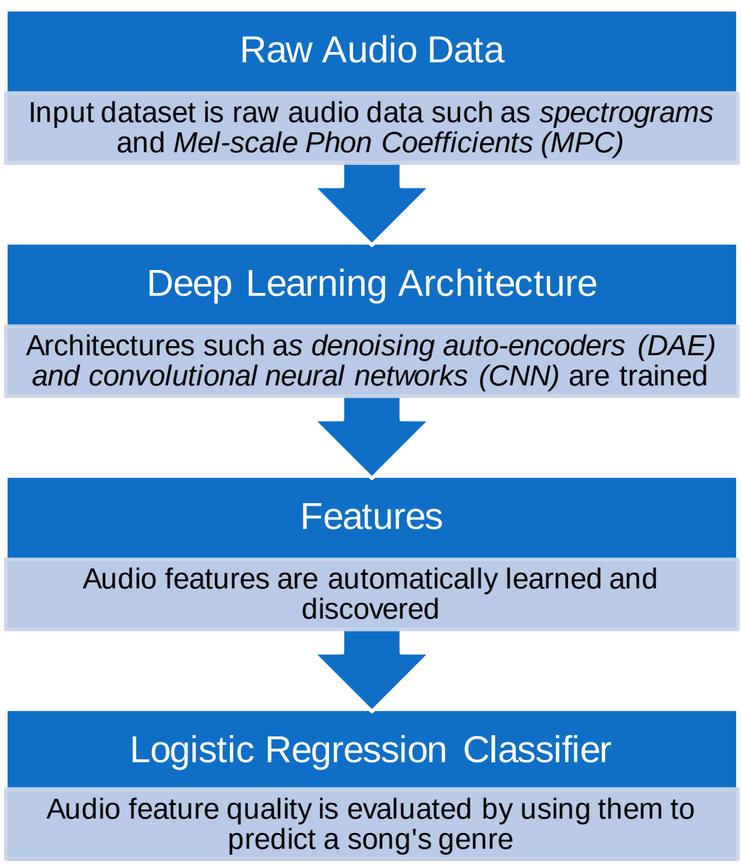



*Fig 2: Methodology used to extract and test audio features*

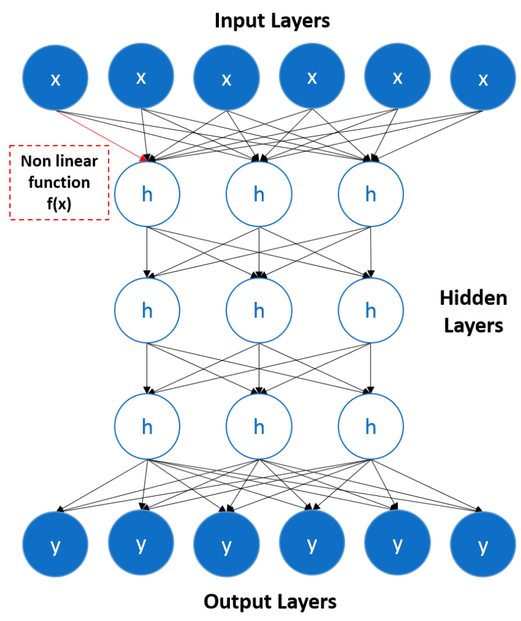



*Fig 3: Network of input, hidden and output units fully connected by non linear functions*

- Denoising auto-encoders were the main deep learning architectures used. These are shown in *Fig 4* below.

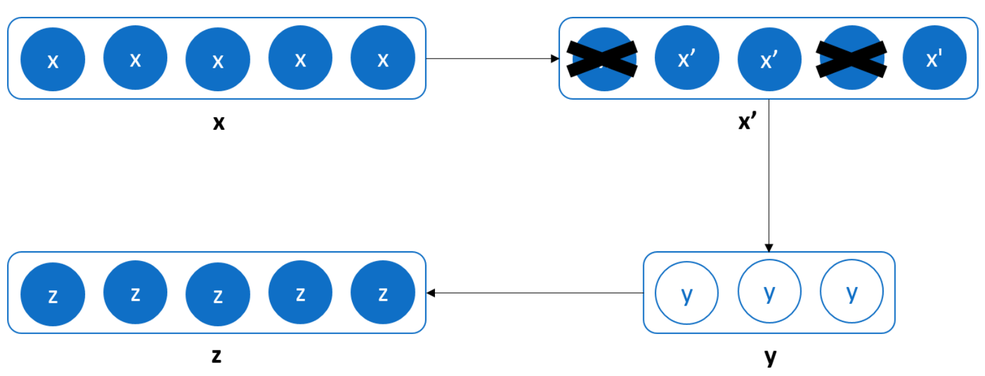



*Fig 4: Input data **x** is corrupted or 'denoised' to **x'**. The auto-encoder maps it to **y** and tries to reconstruct **x**. The reconstruction is outputted as **z**.*

## Results and Discussion

| Dataset | Raw | Learned Features |
|---|---|---|
| **GTZAN Genre** | 61.7% | 67.7% |
| **ISMIR 2004** | 61.1% | 82.6% |

*Table 1: Comparison of classification using raw audio data vs classification using automatically learned* features

- A visualisation of the learned features shows that they are able to capture meaningful structures from the input data.

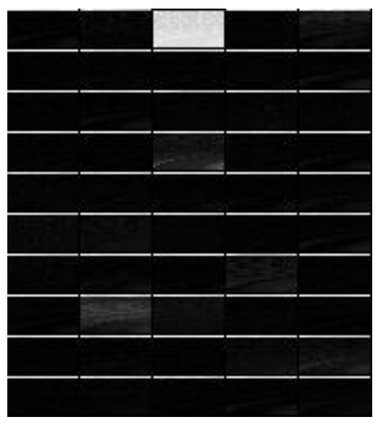


*Fig 5: Feature map of raw input*

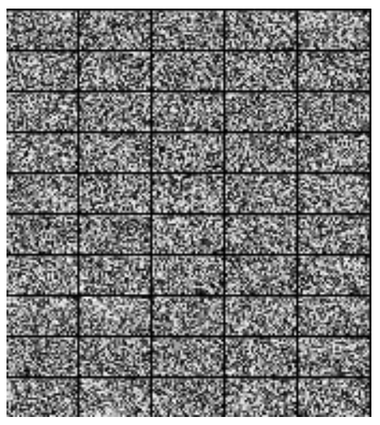


*Fig 6: Feature map at 50% accuracy*

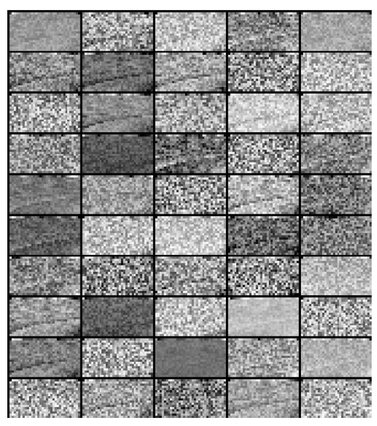


*Fig 7: Feature map at 60% accuracy*

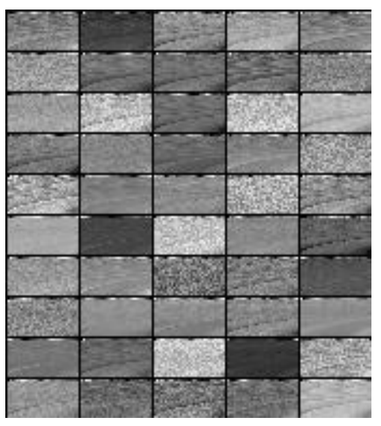
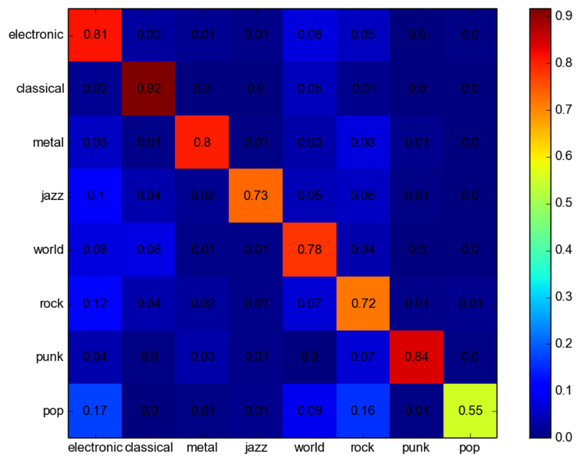


*Fig 8: Feature map at 80% accuracy*

- The feature maps above represent each input value or feature as a grayscale pixel intensity value.
- As the classification accuracy of the features increases, the features become less noisy and more structured.
- Marbled linelike features can be seen in *Fig 7,* becoming more defined in *Fig 8*.




- The confusion matrix shows that more than 70% of data points from all genres but Pop are classified correctly.
- This means the learned features are able to *generalise* all 8 genres of music.

*Fig 9: Confusion matrix showing classification accuracy for each genre*

## Conclusion

- The automatically learned audio features were superior to the raw input data in the genre classification task.
- The genre classification accuracy using the automatically learned features was comparable to the current state of the art results on the ISMIR2004 genre dataset[1].
- The automatically extracted audio features were able to be successfully used in genre classification.

## Future Work

- We will perform artist classification tasks using automatically learned features, moving towards the goal of a music recommendation system based on these features.

## References

[1] Cano, Pedro, et al. "ISMIR 2004 audio description contest." *Music Technology Group of the Universitat Pompeu Fabra, Tech. Rep* (2006).