Can neural networks be used to analyse audio and generate musical notation for a drum kit?

I would like to do my dissertation on neural networks because for some time, I have had a keen interest in AI techniques like NN's and genetic algorithms. I also play musical instruments (drum kit and piano) and have in the past transcribed music by listening to it in segments over and over again.

Although my dissertation will be based on the sounds of a drum kit the implications for a piece of software that can recognise an instrument and transcribe it is huge. If computers can recognise different instruments in a song, then there isn't a reason whole songs, with their individual components, can't have individual transposed sheet music.

After a bit of research, I can see that I'm not the first to have this idea, but recently various programs have started to automate music transcription including:

- https://scorecloud.com/
- https://www.lunaverus.com/

With Lunaverus, the creators claim that they use a large data set of instruments to train a NN to recognise a range of instruments, audio files are then made into sheet music using they're trained NN in a process which takes 3-5 minutes; however, these both seem to find the notes irrespective of what instrument made them and put them into one piece of sheet music.

My dissertation is unique though because I would like to train the neural network to recognise different drums, cymbals and (if I have time) rudiments/techniques (if this is even possible) which don't have a specific pitch, so the NN will just have the pattern of frequencies/amplitudes to go on; because drum skins can be adjusted to change the pitch.

Though this will probably mean that there will be problems recognising drums from different drum kits, but this will be fun to try out and even if it doesn't work generally I should (in theory) still be able to train the NN on how a particular drum kit sounds.

However, some progress has been made in general sound recognition by Google with their Audio library, which at the very least I could attempt to use their extensively labelled audio collection for test data:

- https://research.google.com/audioset/index.html

I have done a bit of research on how audio data is analysed, so that I know where to start, and it seems typical that audio files are converted into an audio spectrogram which can then be analysed by a CNN (convolutional neural network) to train it.

My plan will be (researching as I go):

- To gather as much audio on drum and cymbal sounds as I can find
- If these aren't labelled already label them with what the sound is (in a way the NN will understand)
- Experiment with data augmentation to transform my data to create more labelled data
- Build a CNN which analyses audio spectrograms and produce different versions of this comparing accuracy
- Get a finished CNN and produce a program which takes an audio file and creates sheet music.
- If possible, create a version of this which can transcribe, on the fly, as you play to it.

I still have many questions about this process though, like: how MIDI and audio files work, how best to structure the NN (how many layers yield the highest accuracy) and how will I take some drum data and use this to produce sheet music.