Carnegie Mellon University Africa
Certificate I: Understanding AI and Machine Learning in Africa

Course AIML02: AI and Machine Learning in Africa

Module 02: Application Case Studies
Lecture 01: Healthcare

Welcome to Module 2, Lecture 1, of Understanding AI and Machine Learning in Africa.

In Module 2, we will be looking at six case studies of the application of AI and machine learning in Africa.

For this first lecture, we will focus on a case study in healthcare, looking at the ways that AI can save lives by detecting respiratory distress in newborn babies.

While the remaining five lectures will focus on a single published article, walking through the AI and machine learning techniques that are used to provide the required functionality, here we are going to tell the story of the development of the application, from initial idea to clinical trials, filling in the details of the techniques that are used to make this application successful by studying two technical articles. For that reason, this lecture is longer than usual.

In this lecture, we introduce the first target article for the case study, one that tells the story of how an idea was turned into a successful application, and the people that made this happen.

We then introduce the second target article and walk through the technical details of the machine learning technique that was used to build the applications.

Finally, we walk briefly through a third target article which compares the performance of the original machine learning technique with other techniques that were developed later.

We finish up by summarizing what we have covered and identifying the articles that you should read to consolidate what you have learned.

As we progress through these two articles, we will encounter AI tools and techniques that we mentioned in the first course, AIML01, but you have not yet studied in detail.

As we do so, we will revisit their introduction in AIML01, providing a little more detail if necessary, and we will identify the courses in this and other certificates where these techniques are covered in greater depth.

Once you have listened to this lecture and read the commentary, you should then read the three articles listed at the end: one recounting the birth of the idea, the other two explaining the different AI and machine learning techniques. You should then listen to the lecture again and re-read the commentary

We have three learning objectives, so that, after studying the material covered in this lecture, you should be able to do the following.

1. Explain how AI and machine learning can be used to analyze the cry of a newborn baby and flag the presence of a potential life-threatening condition.

2. Identify the many stages involved in going from an initial idea to a working application.

3. Distinguish the relative strengths and weakness of different AI and machine learning techniques as potential solutions

Slide 1 Welcome to Lecture 1 of Module 2.

In Module 2, we will be looking at six case studies of the application of AI and machine learning in Africa.

For this first lecture, we will focus on a case study in healthcare, looking at the ways that AI can save lives by detecting respiratory distress in newborn babies.

While the remaining five lectures will focus on a single published article, walking through the AI and machine learning techniques that are used to provide the required functionality,

here we are going to tell the story of the development of the application, from initial idea to clinical trials,

filling in the details of the techniques that are used to make this application successful by studying two technical articles.

For that reason, this lecture is longer than usual.

Slide 2 Before we begin, let's set the scene by explaining what the application does.

Ubenwa is a mobile-based app that listens to the cry of a newborn baby and detects early signs of neurological and respiratory conditions.

The specific condition that Ubenwa initially targetted is perinatal asphyxia, a medical condition that deprives the brain of oxygen in newborn babies, resulting in serious complications and possible death.

Ubenwa means “cry of a baby” in the Igbo language of Nigeria.

Slide 3 The company's mission is to the use AI to create accessible tools, for parents, primary care physicians, and researchers.

Slide 4 Let's take a look at the story of Ubenwa.

This part of the case study is based on an article written by Diane Weidner in 2021, entitled: “The Birth of an Idea”.

Ubenwa is the brainchild of Charles Onu, who grew up in Nigeria, and wanted to be a doctor but loved mathematics.

He decided to pursue a BEng degree in Electrical Computer Engineering at the Federal University of Technology in Owerri.

In 2008, he began volunteering for Enactus, a global non-profit organization, first as a member then as president.

That’s how he first learned about perinatal asphyxia.

He explains that “This is a very big problem for resource-limited communities in developing countries like Nigeria where sophisticated detection tools are not available”.

Slide 5 He decided to combine his passion for math and medicine.

His idea was to develop a smart tool to identify babies who are showing signs of respiratory distress by analyzing their cry patterns.

In 2015, Onu joined forces with Innocent Udeogu, a former schoolmate, and they began to work on this problem together

Slide 6 Onu first came to Montreal in 2015 on a Jeanne Sauvé Foundation fellowship as a visiting researcher

He applied to McGill, where he completed a Master’s degree in Computer Science under the supervision of Professor Doina Precup,

During this time, while working on a project with the Montreal Children’s Hospital, he met Latremouille, a Montreal native who had earned her BSc in Physiology at McGill in 2013.

Slide 7 The Ubenwa Team.

Charles Onu is now the Founder and AI Research Lead of Ubenwa.

Innocent Udeogu leads the Software Engineering Development and is the contact person on the ground in Nigeria.

Samantha Latremouille, a PhD student in Experimental Medicine at McGill, is the Clinical Lead.

Slide 8 Ubenwa is a cost-effective and non-invasive tool that allows clinicians to flag the risk of newborn asphyxia, which, if detected early, is easy to treat.

It has the potential to save lives and improve outcomes for millions of newborns each year.

Slide 9 In 2019, Ubenwa was recognized by the World Health Organisation as one of top 30 healthcare innovators emerging from Africa.

Slide 10 Since 2017, to support clinical studies in Nigeria and Canada, Ubenwa has received funding from

Mila – the Quebec Artificial Intelligence Institute,

the Ministère de l'Économie et de l'Innovation (MEI) de Québec, and

the District 3 Innovation Centre.

Slice 11 These clinical studies are a critical step along the road to commercialization.

Slide 12 In 2021, Ubenwa partnered with a top neonatologist in Nigeria to validate its tool for cry-based neurological screening in newborns.

Slide 13 In 2022, Ubenwa secured $2.5 million funding, backed by Radical Ventures and AI luminary Yoshua Bengio.

Slide 14 Now let us turn our attention to the second article in the case study, one of the first technical papers to be published in setting out the machine learning techniques use to solve the problem.

This is a short paper, without too many technical details, so it's a good way to begin our journey through the six application case studies in this module. The amount of technical detail increases as we progress from case study to case study.

As it does, remember that these case studies should be viewed as prequels to motivate the deeper study in later courses.

In 2016, nearly 3 million babies died within 28 days of being born.

Birth asphyxia, has been identified by the World Health Organisation (WHO) as one of the top 3 causes of newborn mortality globally.

It also results to severe, life-long disabilities (such as cerebral palsy, deafness, and intellectual difficulty) in over 1 million infants, annually.

Slide 15 Early detection can prevent this, but many developing countries do not have access to the required diagnostic equipment.

For example, clinical diagnosis of asphyxia typically involves analysis of an arterial blood sample of the infant to measure blood gases, pH, oxygen saturation and electrolytes, using a blood gas analyser.

This information combined with the APGAR score - a standard physical assessment of the newborns based on 5 parameters - gives conclusive evaluation of the presence and/or severity of asphyxia.

Whereas blood gas analysis is a routine procedure for newborns in developed countries, in many developing regions of the world it is not.

Consequently, breathing conditions like asphyxia are generally detected only when the visual symptoms (such as pale/bluish limbs) have emerged, at which point severe neurological damage may have already occurred.

Thus, a high proportion of morbidity and mortality resulting from birth asphyxia occurs in low- and middle-income countries.

.

Slide 16 A diagnostic method for birth asphyxia that

- allows early detection.
- is inexpensive and easy to use.

would allow community health workers, midwives, nurses and even parents to use it worldwide to make life-saving diagnosis of asphyxia.

Slide 17 Ubenwa uses a traditional machine learning approach to classify a baby's cry as normal or asphyxia

Specifically, it uses a support vector machine classifier with features derived from the mel-frequency cepstrum of the baby's cry.

These features are the mel-frequency cepstral coefficients, MFCC.

MFCCs are widely used in automatic speech recognition problems as they provide a representation of audio signals that closely mimic the human auditory system.

Let's walk through the process. The target article assumes the reader is familiar with all the concepts. In case you're not, we'll provide some additional explanations here, as we do in all the case studies.

Slide 18 First, the audio signal of the baby's cry is broken down into time segments.

These are sometimes referred to as signal windows or frames.

The target paper doesn't specify the length of the segment, but the next paper we study does.

In that paper, each segment or frame is 30 ms long

Slide 19 and each segment is extracted every 10 ms.

Slide 20 Thus, the windows and frames overlap by 20 ms.

Slide 21 The next step is to generate the mel-frequency cepstrum for each segment.

The mel-frequency cepstum is a representation of the audio signal that captures information about any periodic structure in the frequency representation of the audio signal, in a manner that reflects the perceptual characteristics of humans (that's the mel part).

There's a lot in that sentence so let's break it down a little.

Slide 22 First of all, we can represent audio signals in two ways.

For simplicity, let's assume with are dealing with digital signals, i.e., discrete, sampled, quantized signals, rather than real-valued analog signals.

Thus, the signal is simply a sequence of samples of the analog audio waveform at discrete time intervals, each sample represented by an integer value, rather than a real value.

This is the first way of representing an audio signal: a sequence of signal samples of the amplitude of the sound as a function of time.

Slide 23 The second way is to represent explicitly the many different frequencies in the signal, i.e, the sound.

The unit of frequency is the hertz, abbreviated Hz. A 1 Hz signal has one cycle per second, e.g., a sine wave with a period of one second. A 2 Hz signal has two cycles per second, e.g., a sine wave with a period of one half a second.

There is an inverse relationship between the frequency and the period of the signal.

Think of period as the time it takes for the signal to repeat: the higher the frequency, the shorter the time it takes for the periodic signal to repeat.

A 1000 Hz, or 1kHz, signal repeats 1000 times each second.

This second, frequency-based, representation is often referred to as the frequency power spectrum. It gives the power, or amount, of each frequency in the original time-domain signal.

Slide 24 The frequency power spectrum representation is generated using the Fourier Transform, for example, using one of many algorithms, such as the Fast Fourier Transform.

Slide 25 The next step in generating the mel-frequency cepstrum is to remap the frequencies so that their spacing is more aligned with human perception of equal changes in pitch.

This is the mel step. Mel comes from melody.

Slide 26 The next step in generating the mel-frequency cepstrum is to take the log of the magnitude of these remapped frequencies.

Slide 27 Finally, in order to reveal the periodicity in the frequency spectrum, we transform the log magnitude representation using a variant of the Fourier Transform: the Discrete Cosine Transform.

Slide 28 The amplitudes of each of the samples in this final representation are the mel-frequency cepstral coefficients: MFCCs.

Again, the target paper doesn't specify the number of coefficients, but the next paper we study does.

In that paper, each frame produces 40 MFCCs.

A one-second (or 1000 millisecond) audio signal of the baby's cry is analysed. Since the frames are shifted by 10 ms as the signal is windowed, there are 101 sets of MFCCs.

These are concatenated to produce a 2-D array representation comprising 40 x 101 elements: 40 MFCCs for each frame in the signal.

These are the features that will be the inputs for the support vector machine classifier, one set of MFCCs at a time.

Slide 29 We briefly met support vector machines in AIML01, Module 2, Lecture 3 on statistical machine learning.

We cover them in detail in AIML05 Essentials of Practical Machine Learning.

Compared to other classifiers like neural networks, SVMs work effectively with limited number examples and high-dimensional data, as is the case here.

SVMs are designed to work effectively with limited examples and high-dimensional data, as is the case in the classification problem here with 40 MFCCs.

Slide 30 The system was trained and tested using the Baby Chillanto Database from the National Institute of Astrophysics and Optical Electronics, Mexico.

The database contains cries of 69 normal, asphyxiating, and deaf infants.

Deafness is one of the most common disabilities resulting from asphyxia.

1389 samples of normal and asphyxiating cries were extracted, with 80% of the data was used to train the classifier and 20% for the test set.

Slide 31 A sample is classified as normal or *asphyxia*, if the majority of its frames, that is, segment, were classified as such.

Slide 32 Results showed sensitivity, that is, accuracy in detecting asphyxiating infants, of 85%

and specificity, that is, accuracy in detecting normal infants, of 89%.

Slide 33 The authors of the target article acknowledge that machine learning is only a part of the solution.

For deployment, they built a mobile application based on the SVN classifier.

naming it *Ubenwa* which, as we already noted, means "cry of a baby” in Igbo language of Nigeria.

Validation is planned by carrying out data acquisition at two hospitals: University of Port Harcourt Teaching Hospital (UPTH), Port Harcourt, Nigeria and McGill University Health Centre (MUHC), Montreal, Canada.

Slide 34 Future enhancements include

Training for robustness to noise in the environment

Identifying the shortest feasible recording length for accurate diagnosis

Optimising the implementation to reduce memory and computational requirements

The target article also mentions the need to investigate alternative machine learning techniques, such as recurrent neural network, including long short-term memory LSTM models.

We briefly met recurrent neural networks and long memory architectures in AIML01, Module 2, Lecture 2 on connectionist approaches to AI.

We cover them in detail in AIML11 Recurrent Neural Networks.

Slide 35 We now turn our attention to the third article in the case study, which Charles Onu subsequently published in 2019 and in which he and his colleagues compared the SVM approach we summarized above with an approach based on transfer learning and convolutional neural networks, ResNet, in particular.

We met convolutional neural networks and ResNet in micro course AIML01, Module 2, Lecture 2 on connectionist approaches to AI

and we mentioned transfer learning in the micro course AIML01, Module 3, Lecture 1: AI Applications in Medicine.

Transfer learning refers to the practice of first training a model, typically a deep neural network model, using a large general-purpose dataset and then tuning the trained model using a smaller application-specific dataset.

We cover both convolutional neural networks and transfer learning in detail in AIML10 Introduction to Deep Learning.

Our goal here is simply to outline this alternative approach as a way of emphasizing their relevance to AI and machine learning in Africa. We won't go into the detailed operation of ResNet, leaving that for micro course AIML10.

Slide 36 The good news is that the mel-frequency cepstral pre-processing is the same as before, with MFCC coefficients being used as the input to the ResNet convolutional neural network.

So, we can skip that part, apart from noting that the complete 2-D MFCC array is provided as input to the ResNet CNN, one array for each baby cry in the data set.

Slide 37 Each convolution layer in the ResNet network has 45 3x3 kernels. The significance of this will become apparent when you study deep neural networks in AIML10. For now, it simply means that ResNet learns 45 individual features.

There is one odd element in the target paper. It states that the ResNet architecture is identical to the res8 architecture in a paper by Tang and Lin (2018). However, the Tang and Lin paper uses three residual blocks in res8 and six in res15.

Slide 38 The final layer is a softmax layer that produces the probabilities of each of the two classes: normal and asphyxia

Slide 39 For transfer learning, ResNet was first pre-trained on three audio data sets

VCTK
Speakers in the Wild (SITW) and
Speech Commands (SC)

These are used for speaker identification, gender classification, and word recognition, respectively.

Slide 40 For each case, ResNet was then post-trained on the Chillanto dataset, as before, with 1049 recordings of normal infants and 340 recordings of infants clinically confirmed to have perinatal axphyxia.

This time, the samples were split into training, validation, and test sets, in the ratio of 60:20:20.

The validation set is used to adjust the ResNet hyperparameters, that is, the parameters that govern the network learning.

We won't go into the details of the training procedure here, but it would be worthwhile revisiting this case study, and the approach taken to training ResNet, once you have completed AIML10 Introduction to Deep Learning and are familiar with deep neural networks, the ResNet architecture, and the training process.

For reference:

Pre-training on the three datasets carried out to obtain performance comparable with the literature, fine-tuning hyperparameters as necessary.

For transfer learning on the target task, models were trained for 50 epochs using stochastic gradient descent with an initial learning rate of 0.001, decreasing to 0.0001 after 15 epochs, a fixed momentum of 0.9, batch size of 50, and hinge loss function. These details will make sense after you have studied deep neural networks in micro course AIML10.

Slide 41 The results show that the best performance was achieved by transfer learning with the speech commands (SC) data set.

However, the original support vector machine approach was the second-best performing model.

For sensitivity, that is, accuracy in detecting asphyxiating infants, SVM achieves 82% and SC Transfer 84%

For specificity, that is, accuracy in detecting normal infants: SVM achieves 87% and SC Transfer 89%

The UAR unweighted average recall figures relate to the the performance on the validation dataset for choosing the best hyperparameter values.

The no-transfer results are for the second baseline (SVM is the first): a ResNet without transfer learning.

The authors also carried out tests to evaluate the performance of the models in the presence of different noise conditions, the sensitivity to the length of the audio recording, varying the length from 0.1 s to 1.0 s, and the sensitivity to different frequency bands by removing bands from the MFCC filter bank.

Slide 42 Result show that in all but one case, the neural models degrade more slowly than SVM in the presence of noise.

Slide 43 The neural models also perform better for shorter recordings, with sc-transfer maintaining peak performance even down to 0.5 s.

Note that in this graph, the length of the recording decreases from left to right; the performance for the full 1 s recording is on the left

The performance of SVM decreases steadily as the duration decreases

Slide 44 sc-tranfer was the most resilient to removal of frequencies

Removing higher frequencies has little impact on performance

To summarize:

1. The use of the infant cry as input for diagnosis of asphyxia presents significant economic, social and clinical benefits.

2. Diagnosis can be performed using techniques based on a support vector machine supervised machine learning, using mel-frequency cepstral coefficients as the feature vector.

3. Performance with a ResNet deep neural network and transfer learning is slightly better.

4. Compared to the current method using a blood gas analyser, Ubenwa is:

   - non-invasive ... there's no need to take a blood sample.
   - low-cost ... just the cost of a smart phone.
   - requires little or no skill to operate.
   - delivers results much quicker, under 20 seconds.

Here are the target articles used for the case study. Please read them carefully.

Onu, C. C., Udeogu, I., Ndiomu, E., Kengni, U., Precup, D., Sant'anna, G. M., Alikor, E. A. D., and Opara, P. (2017) Ubenwa: Cry-based Diagnosis of Birth Asphyxia, Machine Learning for Development Workshop, 31st Conference on Neural Information Processing Systems. https://arxiv.org/pdf/1711.06405.pdf

Onu, C. C., Lebensold, J., Hamilton, W. L., and Precup, D. (2019). Neural Transfer Learning for Cry-based Diagnosis of Perinatal Asphyxia, 20th Annual Conference of the International Speech Communication Association INTERSPEECH, pp. 3053–3057. https://www.isca-speech.org/archive/pdfs/interspeech_2019/onu19_interspeech.pdf

Weidner, D. The Birth of an Idea (2021) https://www.mcgill.ca/channels/article/birth-idea

Here is the reference for the paper that has more details on the ResNet architecture used by the authors.

Tang, R. and Lin, J. (2018). Deep Residual Learning for Small-footprint Keyword Spotting, https://arxiv.org/pdf/1710.10361.pdf