BMJ Open

Efficacy of deep learning methods for predicting under-five mortality in 34 low-income and middle-income countries

Adeyinka Emmanuel Adegbosin [ID],[1] Bela Stantic,[2] Jing Sun [ID] [1]

[1]School of Medicine, Griffith University, Gold Coast, Queensland, Australia
[2]School of Information and Communication Technology, Griffith University, Nathan, Queensland, Australia

Correspondence to
Dr Jing Sun;
j.sun@griffith.edu.au

ABSTRACT

Objectives To explore the efficacy of machine learning (ML) techniques in predicting under-five mortality (U5M) in low-income and middle-income countries (LMICs) and to identify significant predictors of U5M.

Design This is a cross-sectional, proof-of-concept study.

Settings and participants We analysed data from the Demographic and Health Survey. The data were drawn from 34 LMICs, comprising a total of n=1 520 018 children drawn from 956 995 unique households.

Primary and secondary outcome measures The primary outcome measure was U5M; secondary outcome was comparing the efficacy of deep learning algorithms: deep neural network (DNN); convolution neural network (CNN); hybrid CNN-DNN with logistic regression (LR) for the prediction of child's survival.

Results We found that duration of breast feeding, number of antenatal visits, household wealth index, postnatal care and the level of maternal education are some of the most important predictors of U5M. We found that deep learning techniques are superior to LR for the classification of child survival: LR sensitivity=0.47, specificity=0.53; DNN sensitivity=0.69, specificity=0.83; CNN sensitivity=0.68, specificity=0.83; CNN-DNN sensitivity=0.71, specificity=0.83.

Conclusion Our findings provide an understanding of determinants of U5M in LMICs. It also demonstrates that deep learning models are more efficacious than traditional analytical approach.

Strengths and limitations of this study

► The models were tested using a very large data sample, drawn from over 1 million households.
► The survey used a cluster sampling approach and is representative of each country included.
► Socioeconomic, political and cultural differences between the included countries may limit generalisability of the results.
► The cross-sectional design of the study means we can only infer association and not causality.
► Our study does not reflect subnational trends and patterns.

INTRODUCTION

Recent global estimates showed that 5.3 million under-five deaths occurred in 2018; this is equivalent to 15 000 deaths every day and 39 deaths per 1000 live births.[1] A majority of the children who die before their fifth birthday live in sub-Saharan Africa and Southeast Asia; most of these deaths result from preventable and treatable causes.[1 2] Although these estimates represent a significant improvement in under-five mortality (U5M) levels when compared with the levels in the early 1990s, 'preventable death of one child is still too many'.[1 2]

High levels of U5M in low-income and middle-income countries (LMICs) is usually a syndromic feature of a weak health system,[3] and U5MR is a key barometer of the state of a nation's health system and an important impact measure that is reliant on health system input such as health financing, health workforce and infrastructure.[3 4] These inputs in turn determine health service access, readiness, quality and safety and consequently influences coverage of interventions such as antenatal care coverage, postnatal care, demand for family planning satisfied, skilled birth attendance, care for childhood illnesses, nutritional supplementation, etc.[4 5]

Studies have shown that improving child survival requires engaging intricately with a host of child health determinants, including biological, environmental and socioeconomic factors such as level of maternal education, household income, environmental sanitation and hygiene.[5–7] The framework of distal and proximate social, environmental and biological determinants was first described by Mosley and Chen.[5] Unfortunately, many LMICs are constrained by limited finances and limited health budgets, and are unable to intervene on all of the determinants of child health at the same time.[3] It is therefore

increasingly important to identify the most important determinants to be prioritised and to determine the most pressing socioeconomic issues that can serve as a starting point for government and policy makers to focus on intervention strategy.

Furthermore, intervention measures need to be equity-oriented in order to be effective.[8] Hence, disaggregated household level monitoring of coverage and impact indicators are crucial for informing policies and programmatic interventions in the sustainable development goal (SDG) era.[9] It is important to understand the status of every child as against simply exploring global trends, in order to 'leave no one behind' and to 'reach the furthest behind first'.[10] In light of the SDG pledge, monitoring changes at household or community level may require new methodological approaches in engaging with the 'big data', which continues to be generated through ongoing household surveys such as the Demographic Health Survey (DHS) and Multiple Indicator Cluster Survey.[11 12] An expansion of traditional analytical approach may be pertinent and key to effectively monitor health intervention coverage and impact. Machine learning (ML) techniques may represent a novel analytical approach to unravel previously unseen trends; these techniques expand on existing statistical approaches and use methods that are not based on a priori assumptions about the distribution of the data.[13]

Artificial intelligence (AI) described 'as a scientific discipline rooted in mathematics, philosophy and computer science attempts to develop systems with properties of intelligence'.[13] ML is a subdiscipline of AI, 'where computer programs learn to solve new problems for which they weren't explicitly programmed, by learning associations and patterns from example data'.[13] ML deploys a broader set of statistical models than those traditionally used in medicine or public health. Example of such being deep learning models.[13] AI and ML techniques broaden existing statistical models and offer additional tool sets to achieve public health milestones that may not have been previously feasible. For example, ML has been used for real-time surveillance of disease outbreak through social media data mining,[14] AI have been used in large-scale evidence synthesis to guide health promotion and health policy.[15] It is important to state that although AI offers new possibilities for targeted and personalised public health practice, its application must still be guided by social and structural determinants of health; this has also been highlighted by other AI researchers.[16]

In a report recently released by the United States Agency for International Development (USAID) centre for innovation and impact, on the use of AI in global health, AI-enabled population health was identified as one of AI use cases that could have the greatest impact on improving health quality, cost and access in LMICs.[17]

AI-enabled population health encompasses public health surveillance and prediction, population risk management, population health intervention selection and targeting.[17] In this current study, we explored the efficacy of deep learning as a technique for population health surveillance and intervention targeting. Deep learning 'discovers intricate structure in large data sets by using backpropagation algorithm to indicate how a machine should change its internal parameter used to compute representation in each layer from the representation in the previous layer'.[18] Deep learning algorithms have shown excellent performance in genomics, proteomics, drug discovery, speech recognition, visual recognition, object detection and several other domains.[18]

There have been numerous empirical studies on the various applications of ML in hospital settings for prognostication,[19 20] triage[21] and prediction of mortality in the hospital setting.[22] However, application of ML is yet to be demonstrated in population health studies, where it may represent a potential transformative tool.[13] The objective of our study is to fill the gap on application of ML in population health studies, and other previously highlighted gaps. One of the previously highlighted gaps concerns the need to identify the most important determinants of U5MR. To explore these determinants, we employed a data-driven approach by using the random forest algorithm for feature selection, rather than using the traditional hierarchical approach for multivariate analysis, which tends to be highly user-driven and usually involving the development of conceptual frameworks that prejudges the relevance of a limited set of determinants (independent variables).[23] Random forest is an efficient classification and regression algorithm that combines several randomised decision trees and aggregates their predictions. It is especially useful when the number of variables is larger than the number of observations.[24]

The random forest approach allows an unlimited number of variables or determinants to be incorporated into the model. The algorithm automatically tests several hypothesis and selects features that best predicts the outcome, based on information gained from each variable.[20]

Another gap is the need for new ways to gain insights and to unravel previously unseen trends in the prediction of U5M from disaggregated household level data. To fill this gap, we also compared the efficacy of deep learning algorithms: deep neural network (DNN); convolution neural network (CNN); hybrid CNN-DNN with logistic regression (LR) for classifying child survival, and for predicting age of death. We hypothesise that deep learning methods will outperform traditional methods such as LR in the prediction of U5M.

Finally, in this work, we make recommendations on ML implementation, and the new regulatory and ethical considerations for the use of novel ML techniques in public health.[25]

METHODS
Data source and analytical tools
We conducted an analysis on DHS data from 34 LMICs. The DHS is a nationally representative household survey

developed by the USAID in the 1980s.[26] The survey provides data on fertility, family planning, maternal and child health, gender, HIV/AIDS, malaria and nutrition.[27] In total, over 350 surveys have been carried out in over 90 countries.[26] The survey uses a two-stage cluster sampling design, further details about the survey and its design are published elsewhere.[27] Combined multicountry data for this study were obtained from the IPUMS-DHS portal.[28] Combined DHS data were available for a total of 34 LMICs on the IPUMS-DHS database. We used all available data in these countries from 1987 to 2017 (see online supplementary table 1) . Permission to use data for all included countries was granted by the DHS programme. Analysis was conducted using Python software V.3.7. The programming codes used for the various analysis are accessible on Github using the following link: https://github.com/drulna/u5mr_predict

Patient and public involvement
There are no patients involved in this study.

Data preprocessing
Any real-world dataset needs preprocessing to convert it into a representation that can be used to train a model. This can heavily affect the model's performance. This dataset had several irrelevant features, such as IPUMS identifiers created to merge multicountry data. We excluded 14 such features and included 41 features in the final model. Like many census data, the DHS data often contain variables with missing observations. All variables except place of residence (rural/urban) had some level of missingness which range from 5% to 60% of the observation in certain cases, we removed all variables on anthropometric measure due to significant missingness. We performed data preprocessing using the forward-fill approach to replace missing data. There exist multiple strategies that can be deployed to handle missing values,[29] we tested other approaches and tested the models accordingly, only the forward-fill approach was found to provide reproducible and plausible outcomes . 'Forward Fill' strategy involves replacing every missing value with the next real values for each column. This clean and preprocessed data were used for the rest of the analysis.

Variables
Outcome variable
The outcome variable is the risk of death before the age of 5 years, measured as the duration of survival in months from birth.

Independent variables (model features)
The determinants included in the model can be broadly classified into maternal-level determinants, household socioeconomic characteristics and child-level determinants.

Maternal factors
These encompasses maternal behavioural and determinants within the reproductive care continuum, which
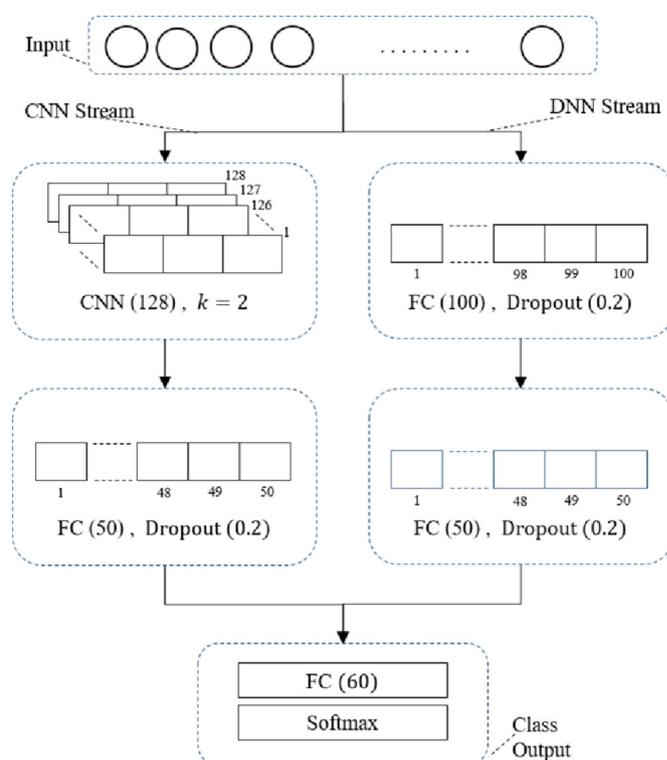
Figure 1 Architecture of the deep neural network (DNN)-convolution neural network (CNN) ensemble model. FC, fully connected.

includes duration of breast feeding, number of antenatal visits when the child was in utero, highest level of maternal education, administration of tetanus injection during pregnancy, provision of prenatal care by a skilled provider, delivery care provider, postpartum health check, unmet need for family planning, prenatal care, pregnancy wanted or not wanted.

Household socioeconomic factors
The household factors included are the household wealth index, the geographical location of the household (urban or rural) and who has final say of the woman's health within the household.

Child-level factors
These include child's postnatal check, sex of the child, oral polio vaccination, measles vaccination, diphtheria, pertussis and tetanus vaccination, BCG vaccination, age of the child and care for childhood illnesses such as diarrhoea and suspected symptoms of pneumonia. Survey-specific definition of all included determinants are published elsewhere.[28]

Feature selection
We use random forest to check feature importance with respect to its predictive power. figure 1 shows the feature importance (red bar) and variance of each tree in random forest (black vertical line). It can be observed that 'duration of breast feeding' has the most importance to predict a child's death. However, there are some features that are of limited importance. We perform feature selection

based on this information. We drop all features whose importance are <0.001, because we found that the accuracy of the classifier does not improve beyond this level, and adding the additional attributes only creates unnecessary additional computational overhead. In total, 29 features fell within our cut-off for feature importance and included in the final model. For comparing the utility of feature selection, we perform two experiments. One without feature selection (on all original 41 features) and one with feature selection (on selected 29 features).

Model selection

We selected multivariate LR as an example of traditional model.[20] Three deep learning techniques (DNN, CNN and DNN-CNN) were selected as modern ML approaches. For all the four models, we pose this problem as a multiclass problem, such that each value in the label is assigned an integer and then we binarize the output (ie, one-hot encoding). All categorical attributes are also converted to numerical, that is, dummy variables, by mapping each unique value to a number. After careful consideration, we concluded that the best ratio for training is 75% of the data, while the remaining 25% of the data are reserved for testing purposes. This choice is in line with literature and close to 80/20, which is quite a commonly used training/ testing ratio, often referred to as the Pareto principle. We compare the performance of LR as a representative of traditional model, with three deep learning methods: DNN, CNN and hybrid CNN-DNN.

Model architecture
Deep neural networks

DNNs are a special kind of neural network with multiple hidden layers and usually hundreds of units in hidden layers. Each neuron of one layer is connected to every neuron of subsequent layer, also called fully connected (FC) layers. For each layer in DNN, a weight matrix is learnt. DNNs act as blackbox and can learn the data representation automatically with backpropagation of the error at final layer. A *softmax* layer is usually used to get final prediction for the class label.

Convolutional neural networks

CNN is a specific deep learning architecture that learns a filter instead of weight matrix. This filter is used to perform convolution with input data to get a feature map. This feature map can then be forwarded to a final *softmax* layer for prediction. A key advantage of using CNN over DNN is that it requires fewer parameters and less iterations to converge as only last layer is FC.

In our presentation, we give results for DNN, CNN and a hybrid of DNN and CNN. We show that the later gives the most optimal results, leveraging benefits of the two worlds.

Hybrid DNN-CNN ensemble model

In this model, the input is forwarded to two streams, where one represents DNN while the other CNN. As our input is one dimensional (1D), we use 1D CNN. With regard to DNN stream, the input is forwarded to an FC layer with 100 units. For non-linearity, the activation function *ReLU* is used which is defined as $\max(0, x)$. This is followed by a batch normalisation (BN) and dropout (DO) layer to avoid feature co-adaptation. Then a second FC layer with 50 units is used to squash the information, which is again followed by BN and DO layers. The output of this layer is forwarded for concatenation with the output of CNN stream (figure 1).

Regarding the CNN stream, the input is forwarded to a 1D CNN layer with 128 filters and kernel/filter size of 2, with *ReLU* non-linear activations. The output is followed by a maxpooling to drop low information activations. This is followed by a BN and DO layer. The information is squashed into an FC layer with 50 units, which is again followed by BN and DO. Finally, the output is forwarded for concatenation with the output of DNN stream.

The combined features of both streams are then forwarded to a single FC layer with *softmax* activation, which results in class probabilities. The class label is assigned based on maximum probability. The detailed diagram of the architecture is shown as figure 1.

To optimise the hyperparameters such as optimizer, DO rate and learning rate, we used grid search. The available choices for hyperparameters and the selected value are given in table 1. To stop the training, we employed early stopping strategy where the training was stopped if the validation accuracy did not improve for 20 epochs. A checkpoint was created at the epoch where the validation accuracy showed improvement as compared with previous checkpoint. The choice of number of layers, number of neurons in each layer and number of filters in CNN was made empirically.

Model evaluation

We evaluated the performance of each model using a receiver operating characteristic (ROC) plot, we also derived the weighted precision, sensitivity (also known as recall), specificity, f1-score and area under the curve

Table 1 Hyperparameters choice and selected values through grid search

Hyperparemeter	Available choice	Selected choice
Optimizer	Adam, Adadelta, RMSprop	Adam
Dropout rate	0.1, 0.2, 0.3, 0.4, 0.5	0.2
Learning rate	0.002, 0.004, 0.006	0.004

(AUC) for each model. The formula for calculating the performance metrics are as follows:

$$\text{Precision} = \frac{True\ Positives}{True\ Positives+False\ Positives};$$ $$\text{F1-score} = 2 \times \frac{precision \times recall}{precision + recall}, \text{ where } recall = \frac{True\ Positives}{True\ Positives+False\ Negatives};$$ $$\text{Specificity} = \frac{True\ Negatives}{True\ Negatives + False\ Positives}$$

The models were evaluated before and after feature selection. Analysis was initially conducted using all preselected variables. We thereafter optimised the various models based on empirical results from the random forest analysis. As this is a multiclass problem, the ROC plots and performance metrics are all based on micro-averages.

RESULTS
Characteristics of the study population
A total population size of (n=1, 520 018) children drawn from 956995 unique households were included in the study, the sample was drawn across 34 LMICs. The sample size drawn from each of the included countries is presented in online supplementary table 1. The mean age of the total children population is 1.89 (±1.40). Majority (n=1 100 211; 72.7 %) resides in rural areas. Just under half (n=636882; 45.2%) were in the lowest two wealth quintile (Q1 and Q2). Majority (n=1 100 262; 73.2%) were uneducated or had only primary education, majority received some form of postnatal check, delivery care tetanus injection before birth and approximately two-third breast fed their children for >6 months (table 2). A total of n=111907 (7.3%) under-five deaths were recorded survey-wide across all 34 countries. Nearly half, 48.9% (n=54825) of these deaths were neonatal death.

Feature importance
Overall, key determinants of U5MR include maternal factors such as duration of breast feeding, number of antenatal visits when the child was in utero, provision of maternal postnatal care by a skilled provider, highest level of maternal education, administration of tetanus injection during pregnancy, prenatal care provision by a skilled provider. Significant household socioeconomic factors include household wealth index and geographical location of the household. Time to child's postnatal check was found to be the most significant child level determinant (figure 2).

Model comparisons (before feature selection)
Comparison of the performance of the models before feature selection reveals that hybrid of CNN-DNN performs the best in terms of all metrics (sensitivity=0.68, specificity=0.83), while LR performs the worst (sensitivity=0.47, specificity=0.53) (table 3).

Figure 3 shows the ROC curves for all the classifiers. It shows that hybrid CNN-DNN model outperforms all other models.

Model comparisons (after feature selection)
We found that feature selection does not improve the performance of LR. However, for all deep learning-based

Table 2 Descriptive analysis of the study population

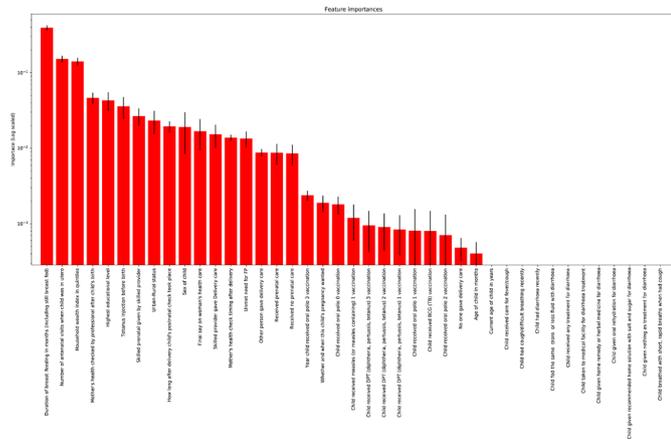| Variables	n (%)	M (SD)
Age of child (years)		1.89 (1.40)
Gender		
Male	775957 (51.0)	
Female	744061 (49.0)	
Residence		
Urban	413705 (27.3)	
Rural	1100211 (72.7)	
Wealth quintile		
Poorest	334135 (23.7)	
Poorer	302747 (21.5)	
Middle	283295 (20.9)	
Richer	260049 (18.4)	
Richest	231177 (16.4)	
Maternal education level		
No education	670115 (44.6)	
Primary	430147 (28.6)	
Secondary	342889 (22.8)	
Higher	60517 (4)	
Postnatal check received		
Yes	283683 (52.6)	
No	255752 (47.4)	
Prenatal care received		
Yes	839821 (79)	
No	223913 (21)	
Delivery care received		
Yes	1393930 (95.8)	
No	60774 (4.2)	
Tetanus injection before birth		
Yes	387999 (75.5)	
No	125889 (24.5)	
Duration of breast feeding		
<6 months	517671 (36.4)	
6 months or more	903170 (63.6)	

N=1520018.

models, feature selection results in performance gain. The most performance gain is shown by CNN-DNN (sensitivity=0.71, specificity=0.83). CNN-DNN model performs the best out of all classifiers in both settings, that is, before feature selection and after feature selection (table 4).

In figure 4, we present ROC curves for all the classifiers. It shows that hybrid CNN-DNN model remains the top performer of all the models.

DISCUSSION
A number of maternal-level, child-level and socioeconomic indicators were found to influence U5M. Duration

Figure 2 Feature importance using random forest.

Figure 3: Micro Average ROC Curve (Before Feature Selection)

Figure 3 Micro-average receiver operating characteristic (ROC) curve before feature selection. CNN, convolution neural network; DNN, deep neural network; LR, logistic regression.
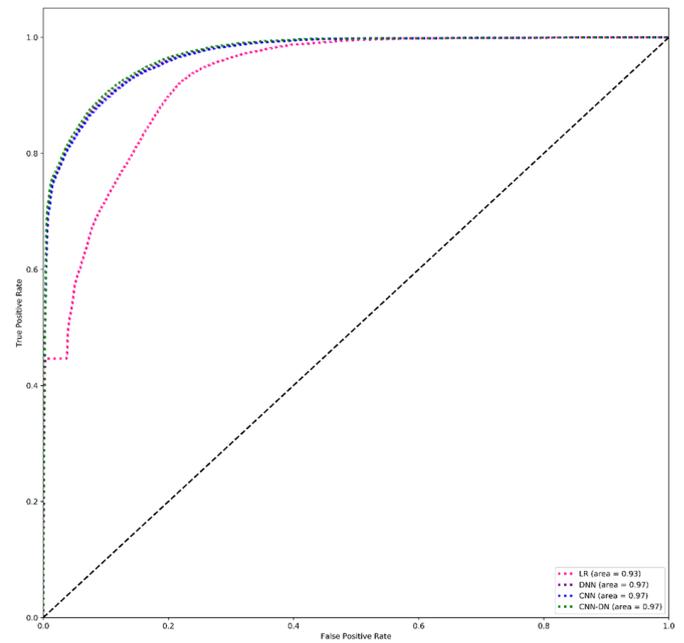
of breast feeding was found to be a significant maternal-level determinant. Previous studies corroborate our findings, it has been shown that children breast fed for a longer duration have lower infectious disease morbidity and mortality, and better chance of survival than those who are breast fed for shorter periods, or not breast fed at all.[30] Multiple studies have also shown that early initiation of breast feeding, and exclusive breast feeding reduces both neonatal and early infant mortality.[30 31] In addition to breast feeding, several other factors within the continuum of essential obstetric care, such as antenatal care visits, postnatal care, delivery care and maternal tetanus immunisation were found to be significant predictors of U5M. These may partly be explained by our finding, which showed that nearly half of the mortality occurred during early neonatal life, which is in line with other previous studies.[32 33] Several previous studies have shown that provision of essential obstetric care is vital for survival during the neonatal period.[34 35] In addition, we found that the household wealth index was a slightly more important determinant compared with maternal level of education. This finding however contradicts the work of Fuchs *et al*, where they argued that mother's education is the fundamental determinant of child mortality and is relatively more important than income level. They argued that education impacts the child's health through better maternal health, increased health-specific

knowledge, avoidance of traditional, harmful behaviours, greater economic resource as a consequence of education and general female empowerment.[36] They however highlighted that other social scientists have often considered education and income as generally highly correlated and tend to be regarded as interchangeable indicators of socioeconomic status.[36]

The timing of the child's postnatal check and the gender of the child were also found to be predictive of child's survival. Postnatal check within 24 hours of birth have been shown to be crucial in identifying, managing or referring complications and ultimately in preventing child mortality.[35]

Our findings regarding the superiority of ML over traditional approaches such as LR in predictive analysis are also in line with findings elsewhere.[20 37]

Table 3 Performance comparison (without feature selection)

Performance metrics	LR	DNN	CNN	CNN-DNN
Sensitivity	0.47	0.67	0.66	0.68
Specificity	0.53	0.84	0.83	0.83
Precision	0.35	0.58	0.57	0.62
F1-score	0.38	0.62	0.60	0.63
AUC	0.93	0.97	0.97	0.97

AUC, area under the curve; CNN, convolution neural network; DNN, deep neural network; LR, logistic regression.

Table 4 Metrics comparison after feature selection

Performance metrics	LR	DNN	CNN	CNN-DNN
Sensitivity	0.47	0.69	0.68	0.71
Specificity	0.53	0.83	0.83	0.83
Precision	0.35	0.62	0.62	0.67
F1-score	0.38	0.63	0.62	0.67
AUC	0.93	0.97	0.97	0.97

AUC, area under the curve; CNN, convolution neural network; DNN, deep neural network; LR, logistic regression.

Figure 4: Micro Average ROC after Feature Selection
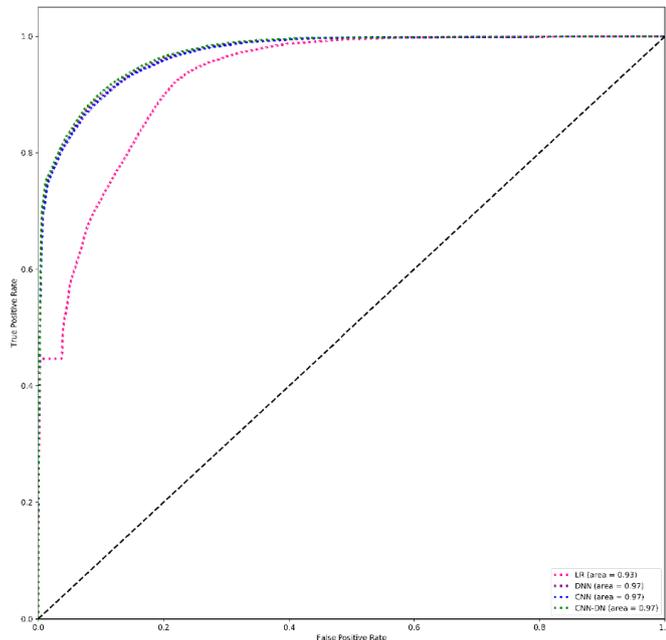
Figure 4 Micro-average receiver operating characteristic (ROC) curve after feature selection. CNN, convolution neural network; DNN, deep neural network; LR, logistic regression.

This study however has some limitations. First, this is proof-of-concept cross-sectional study; hence, we can only draw inference on associations, and not on causality. Second, we did not measure change over time. Future studies should consider incorporating temporal data points, to draw inference on changes over time, and possibly causality. Finally, we did not explore individual country, regional and subgroup level variations and cannot conclude that the degree of association is the same across different countries and subgroups, due to differences in socioeconomic, geographical, cultural and political realities. Hence, future studies should consider disaggregating with stratifiers such as income, education and place of residence, to explore subgroup differences.

Recommendations for ML implementation, governance and ethics

Our recommendations regarding the implementation and regulation of ML are fourfold. First, there is a burgeoning risk that the adoption and benefits of ML may be imbalanced.[38] High-income countries are beginning to increasingly adopt and benefit from deploying some of these novel technologies; therefore, there is the risk of extending the disparity between LMICs and high-income countries even further. To achieve equity in the implementation of this technology, there is a need for capacity building across board and collaborative use of technological resources between LMICs.

Second, regarding AI research governance and ethics (regulation), the capabilities of AI application in public health are not yet fully understood, and its application is still evolving. This implies that any regulatory attempt will effectively require understanding the capabilities of AI as a tool in public health and medicine. Like other medical research endeavours, the regulatory framework and ethical guidelines will have to evolve, as our understanding of the application of AI evolves. As such, we posit that there is a concordance between regulation, governance, research and development of AI technology. In the light of this, we suggest collaboration between research institutions, academic stakeholders, policy makers and regulatory authorities. There is a need to engage with all stakeholders across the spectrum of AI research, development and ethics.

Third, we believe that existing medical research ethical guidelines are highly applicable and cover several aspects of ML research. However, there is a need to strengthen regulatory aspects pertaining to data security and protection. The growth in the adoption of ML analytical techniques will usher an increase in the level of data transactions and with this, comes the potential risk of breaches to health data privacy. There are existing capabilities to re-identify anonymised data, using a few parameters within the data. Hence, regulatory efforts need to focus on data security, especially reducing the risks of data re-identification.

Fourth, as knowledge and application of AI continues to grow in leaps and bounds, and while regulatory efforts are still rudimentary and trying to catch up, we envisage a vacuum in governance, which will have to be filled. As such, there may be a need for the development and ratification of regulatory framework, which may be possible through the collaboration of multiple stakeholders.

CONCLUSIONS

This study demonstrates the superiority of ML as a tool for understanding previously unseen insights in large global health data. We have shown that ML algorithms such as random forest, may be more insightful than the user-dependent traditional hierarchical approach of testing a limited set of determinants for outcome prediction in multivariate analysis. Using random forest, we found that duration of breast feeding, household wealth index and level of maternal education are the most important determinants of U5MR. In addition, we also show that deep learning algorithms are more sensitive and specific for the prediction of U5MR and this finding may be applicable to other multivariate models, for data-rich population studies.

Going forward, the most important implication of this study is that if deep learning algorithms such as the one we describe in this study, are deployed in production in combination with spatial data, it is possible to identify and flag children who are most at risk and not likely to survive until the age of 5, such that necessary interventions can be targeted to communities where those children live. To the best of our knowledge, there are no existing studies

that have investigated U5M, using a similar analytical approach.

ORCID iDs
Adeyinka Emmanuel Adegbosin http://orcid.org/0000-0001-5218-1008
Jing Sun http://orcid.org/0000-0002-0097-2438

REFERENCES

1 World Health Organization. Children: reducing mortality [Internet]. Facts sheet children: reduing mortality, 2018. Available: https://www.who.int/news-room/fact-sheets/detail/children-reducing-mortality [Accessed 5 Apr 2019].

2 UNICEF. One is too many Ending child deaths from pneumonia and diarrhoea [Internet]. UNICEF, 2016. Available: http://data.unicef.org/topic/child-health/pneumonia/ [Accessed 5 Apr 2019].

3 Maruthappu M, Ng KYB, Williams C, et al. Government health care spending and child mortality. *Pediatrics* 2015;135:e887–94.

4 World Health Organization. *Indicator and monitoring framework for the global strategy for women's, children's and adolescents' health (2016-2030)*, 2016.

5 Mosley WH, Chen LC. An analytical framework for the study of child survival in developing countries. *Popul Dev Rev* 1984;10:25.

6 Feng XL, Theodoratou E, Liu L, et al. Social, economic, political and health system and program determinants of child mortality reduction in China between 1990 and 2006: a systematic analysis. *J Glob Health* 2012;2:010405.

7 Barros FC, Victora CG, Scherpbier R, et al. Socioeconomic inequities in the health and nutrition of children in low/middle income countries. *Rev Saude Publica* 2010;44:1–16.

8 Adegbosin AE, Zhou H, Wang S, et al. Systematic review and meta-analysis of the association between dimensions of inequality and a selection of indicators of reproductive, maternal, newborn and child health (RMNCH). *J Glob Health* 2019;9:010429.

9 Health Organization W. *Handbook on health inequality monitoring: with a special focus on low- and middle-income countries*.

10 UN department of economic and social affairs. Leaving no one behind | UN DESA | United Nations Department of Economic and Social Affairs [Internet]. Available: https://www.un.org/development/desa/en/news/sustainable/leaving-no-one-behind.html [Accessed 8 Apr 2019].

11 Corsi DJ, Neuman M, Finlay JE, et al. Demographic and health surveys: a profile. *Int J Epidemiol* 2012;41:1602–13.

12 Unicef, Bng-, -Noorani. Monitoring the situation of children and women for 20 years.

13 Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. *J Glob Health* 2018;8:020303.

14 Şerban O, Thapen N, Maginnis B, et al. Real-Time processing of social media with sentinel: a syndromic surveillance system incorporating deep learning for health classification. *Inf Process Manag* 2019;56:1166–84.

15 Michie S, Thomas J, Johnston M, et al. The human Behaviour-Change project: harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation. *Implement Sci* 2017;12:121.

16 Panch T, Pearson-Stuttard J, Greaves F, et al. Artificial intelligence: opportunities and risks for public health. *Lancet Digit Health* 2019;1:e13–14.

17 USAID's Center for Innovation AND Impact. Artificial intelligence in global health defining a collective path forward.

18 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–44.

19 Weng SF, Reps J, Kai J, et al. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One* 2017;12:e0174944.

20 Taylor RA, Pare JR, Venkatesh AK, et al. Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data-driven, machine learning approach. *Acad Emerg Med* 2016;23:269–78.

21 Horng S, Sontag D, Halpern Y, et al. Undefined. creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning, 2017. Available: journals.plos.org

22 Arya R, Wei G, McCoy JV, et al. Decreasing length of stay in the emergency department with a split emergency severity index 3 patient flow model. *Acad Emerg Med* 2013;20:1171–9.

23 Victora CG, Huttly SR, Fuchs SC, et al. The role of conceptual frameworks in epidemiological analysis: a hierarchical approach. *Int J Epidemiol* 1997;26:224–7.

24 Biau G, Scornet E. A random forest guided tour. *Test* 2016;25:197–227.

25 Zandi D, Reis A, Vayena E, et al. New ethical challenges of digital technologies, machine learning and artificial intelligence in public health: a call for papers. *Bull World Health Organ* 2019;97:2.

26 Croft TN, Marshall AMJ, Allen CK. Guide to DHS statistics 2018.

27 Burgert CR and DP. *Linking DHS household and spa facility surveys: data considerations and Geospatial methods. DHS spatial analysis report*, 2014.

28 Boyle EH, King M, MS. *IPUMS Demographic and Health Surveys: Version 6 [dataset]. IPUMS and ICF, 2018*, 2018.

29 Mckinney W. *Pandas: a foundational python library for data analysis and statistics*.

30 Victora CG, Bahl R, Barros AJD, et al. Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect. *Lancet* 2016;387:475–90.

31 Health NSG-TLG. *Undefined. timing of initiation, patterns of breastfeeding, and infant survival: prospective analysis of pooled data from three randomised trials*. Elsevier, 2016.

32 Sankar MJ, Natarajan CK, Das RR, et al. When do newborns die? A systematic review of timing of overall and cause-specific neonatal deaths in developing countries. *J Perinatol* 2016;36 Suppl 1:S1–11.

33 Hug L, Alexander M, You D, et al. National, regional, and global levels and trends in neonatal mortality between 1990 and 2017, with scenario-based projections to 2030: a systematic analysis. *Lancet Glob Health* 2019;7:e710–20.

34 de Souza S, Duim E, Nampo FK. Determinants of neonatal mortality in the largest international border of Brazil: a case-control study. *BMC Public Health* 2019;19:1304.

35 Langlois Étienne V, Miszkurka M, Zunzunegui MV, et al. Inequities in postnatal care in low- and middle-income countries: a systematic review and meta-analysis. *Bull World Health Organ* 2015;93:259–70.

36 Fuchs R, Pamuk E, Lutz W, et al. Education or wealth: which matters more for reducing child mortality in developing countries? *Vienna Yearbook of Population Research.* 2010;8:175–99.

37 Panesar SS, D'Souza RN, Yeh F-C, et al. Machine learning versus logistic regression methods for 2-year mortality prognostication in a small heterogeneous glioma database. *bioRxiv* 2018;472555.

38 Schwab K. The Fourth Industrial Revolution, by Klaus Schwab | World Economic Forum [Internet]. Available: https://www.weforum.org/about/the-fourth-industrial-revolution-by-klaus-schwab [Accessed 5 May 2019].