Hip Fracture Detection Using Artificial Intelligence: A Pilot Study

Samuel Arsenio M. Grozman, MD, Patricio E. Dumlao III, MD and Lauro T. Gonzales, MD

Department of Orthopedics, Philippine General Hospital, University of the Philippines Manila

ABSTRACT

Background. Hip fractures are commonly missed on the first radiograph in up to 30% of patients. The delay in diagnosis leads to significant gaps in management and consequent morbidities. Thus, a computer-aided hip fracture recognition through the Artificial Neural Network deep learning model, which allows the program to learn and gain experience with more images processed, has been created.

The study aimed to determine the accuracy and sensitivity of the artificial neural network model in detecting fractures of the hip and explored the feasibility of its use as a diagnostic screening tool.

Materials and Methods. A sample size of 45 participants/samples per treatment group was computed using a confidence level of 90%, and prevalence of 0.05 for a pilot study. The program was tested by processing digital pictures of radiographs of patients with known hip fractures that included femoral neck, intertrochanteric, subtrochanteric, and proximal femur fractures taken from the database of adult patients, who have undergone surgery for a hip fracture at the Philippine General Hospital from 2016-17. The 90 (45 fractured, 45 normal) manually selected proximal femur images were run on 10 models. The models were based on AlexNet and VGG-16, which are the representative convoluted neural networks designed for image analysis.

Results and Conclusion. The program had an accuracy of 70%, specificity of 42.2% and sensitivity of 97.8%. This study is proof of concept that a deep learning model fracture detection software shows potential in hip fracture detection. Further training is necessary to make this promising innovation clinically useful.

Key Words: artificial intelligence, hip fractures, machine learning

INTRODUCTION

Hip fractures are an increasing source of morbidity and subsequent mortality. Historically, the incidence of hip fractures is bimodal, occurring in young active patients involved in high-energy vehicular crash, and among elderly patients. Hip fractures rise incrementally with age in both sexes.[1] In the two age groups, diagnosis and consequent management is imperative to avert complications associated with delayed management such as deformity, discomfort, pain, and fatal embolic events from prolonged immobilization.[2] For the younger age group, the identification of hip fracture is delayed in up to 31% of patients. This may be due to vertical, minimal-to-nondisplaced fracture lines, with incidence of up to 59% of affected individuals with other ipsilateral injuries.[3] For the older age group with significant decreased bone mineral density, hip fractures are related to low-energy falls such as stumbling from standing height.[1] Though only 1% of these falls will lead to a hip fracture, a missed occult fracture line will result in a 20% increased risk of medical-related morbidity.[2]

Corresponding author: Samuel Arsenio M. Grozman, MD
Department of Orthopedics
Philippine General Hospital
University of the Philippines Manila
Taft Avenue, Ermita, Manila 1000, Philippines
Email: smgrozman@up.edu.ph

To assist physicians in image interpretation, the concept of computer-aided detection was developed.[4,5] In this method, images are run through a programmed algorithm that will permit the system to arrive at a diagnosis. This diagnosis is dependent on the parameters programmed for a specific image of the body part. The concept has been utilized in breast mammography. The program would flag images that satisfied a certain diagnostic criterion as positive. This prompts an alert for the interpreting physician to focus more on the suspected pathologic finding. Other such systems include chest CT and chest radiographs.[4,5] With the advancing development of computer-aided detection, principles of machine learning (ML) were postulated. ML is a concept of how computer systems can summarize patterns from actual data and produce models connecting covariates to a specific point of interest.[6] The concept can thus predict an outcome (e.g., diagnosis, differential diagnosis, and risk) based on empiric information (e.g., registries, medical records, radiographs) fed to the system to aid physicians in making effective and correct judgments.[7]

The concept of computer-aided diagnosis is evolving in developing countries such as the Philippines for applications including telemedicine/telehealth.[8,9] Telehealth systems could employ computer programs that would assist general practitioners in rural and underserved areas in diagnosis, and subsequently, in deciding whether specialist referrals are warranted. Thereby, a study was conducted that determined the accuracy of an artificial intelligence program in diagnosing hip fractures and assessed the feasibility of using the program as a diagnostic screening tool.

METHODS

A deep learning computer program was developed utilizing an artificial neural network model. The program learned to distinguish between fractured and non-fractured hips using 200 anonymized stock digital radiographs of fractured hips and 200 non-fractured hips. This study focused on testing the accuracy of the developed program.

Pilot population size computation and distribution

We computed for the sample size for a pilot study using the formula by Viechtbauer et al. A level of confidence of 90% (90-95%, medically adequate), and prevalence of 0.05 were used to yield a total of 45 participants/samples per treatment group.[10] The radiographs were selected from a database of adult patients who have undergone surgery for a hip fracture at the Philippine General Hospital from 2016-2017.

Radiograph preparation, fracture selection and categorization

We retrieved radiographs of patients with the diagnosis of a hip fracture that was confirmed intraoperatively. Hip radiographs were defined as images showing the proximal portion of the femur, including the entire acetabulum and the femoral head to the sub trochanteric area 1 cm below the lesser trochanter within the field of view. Femoral neck, Intertrochanteric and sub trochanteric fractures were all broadly classified as hip fractures for the purposes of the study. The contralateral, non-fractured, non-operated hip was also run through the program as the control group. Specific inclusion criteria were as follows:

1. Adult patients aged 18 years and older
2. Patients confirmed to have a hip fracture through intraoperative findings
3. Patients with traumatic, fragility or pathologic hip fractures
4. Patients with concomitant injuries to the acetabulum and pelvis
5. Hip radiographs of fractures taken from between one day to one year post injury

Development of the algorithm, schematic, and statistical analysis

The program was tested by processing digital pictures of radiographs of patients with known hip fractures. Stratification as to the different types of fractures were not done in this learning model. Manually selected proximal femur images were run on 10 models. The models were based on AlexNet, and VGG-16, which are both examples of deep learning convoluted neural networks (CNN) designed for pattern recognition. The running time of an AlexNet model is 14 seconds, and 16 seconds for VGG-16 (Figure 1).

Radiographs run through the system were classified as either fractured or normal, based on the program's algorithm. Standard sensitivity and specificity analysis using the standard 2x2 computation table was done. True positives (TP) were defined as radiographs of hips with fractures correctly identified by the program to have a fracture. False positives (FP) were radiographs without fractures, recognized by the program as with fractures. True negatives (TN) were images without fractures detected by the program to have no fractures; and false negatives (FN) were images of fractures, categorized as not having any fractures by the program. Sensitivity was computed as TP/(TP+FN), and specificity computed as TN/(TN+FP). Program accuracy was computed as (TP+TN)/(TP+TN+FP+FN).

The only area of interest identified by the program was the hip. Other concomitant fractures within the images used were outside the scope of the study.

RESULTS

A total of 90 hip radiograph images were interpreted by the program; 45 were images of fractured hips (22 left, 23 right), and 45 were images of the uninjured contralateral hip from the same radiograph. Images of the fractured and uninjured hip were processed separately (Appendix Table 1).

The program was able to make the correct interpretation in 63 out of 90 images, with a 70% accuracy rate. Program specificity was 42.2% and sensitivity was 97.8% (Table 1).
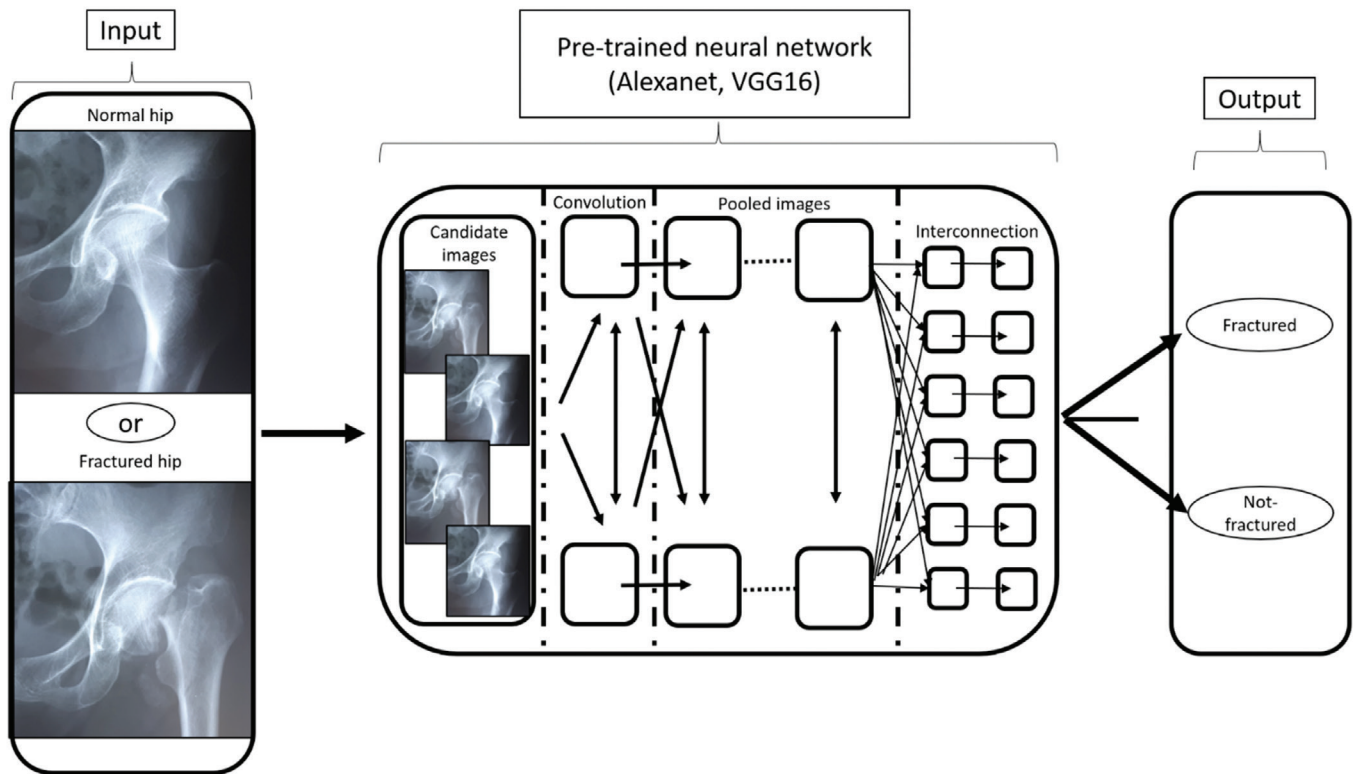
Figure 1. Schematic workflow with representative radiograph.

Table 1. Analysis table

AI interpretation / Radiographs	Fractured	Normal	Total
Diagnosed fractured	44 (TP)	26 (FP)	70
Diagnosed normal	1 (FN)	19 (TN)	20
Total	45	45	90

TN – True Negative; TP – True Positive; FN – False Negative; FP – False Positive

DISCUSSION

An Artificial Neural Network (ANN) model is a computer programming technique that allows computer programs to form associations. It utilizes algorithms it determines as the most appropriate problem-solving tool based on these associations. This is more popularly known as artificial intelligence (AI), or more recently, machine learning (ML). Practical applications for this are abundant: facial and fingerprint recognition of smart phones, smart navigation, and computer games against AI opponents. The main difference between traditional programs versus artificial neural network programs is the ability to form associations and learn.[7,11]

In the field of orthopedics, computer-aided detection of long bone fractures has long been undergoing development. Various algorithms were created, with varying success rates when compared to human physician interpreters. Computers were programmed to detect fractures based on long axis angulation changes, contour edge breaks, image color gradient changes, and even individual image pixel color and grouping. All current published methods of computer-assisted fracture identification, however, rely on a system following a set algorithm to come up with a diagnosis. These programs are designed to target long bone fractures, and the latest developed programs have a detection accuracy of up to 95%.[8,10,12,13]

Two prior programs have targeted hip fractures, with an accuracy of 62% and 85%.[12] The first program used edge detection algorithms to extract the image outline and determine the femoral neck angle, which was used as the basis for detecting fractures. The program used edge detection combined with a Gabor Orientation map algorithm. This used the dominant orientation of the trabeculae taken from sampling locations within the femoral head, and deviation was correlated with a fracture. The femoral neck angle was still used as the basis for detection. These two prior programs were also highly operator-dependent since initialization points for the active contour and sampling areas for the femoral head needed to be manually set.[12] A limitation to the algorithms used in the prior programs is that femur rotation can affect the apparent neck angle.

It is important to emphasize that the previous programs, with static nature of the process, is opposed to ML from the artificial neural network that was utilized in this study. The ML program was able to learn and improve through time

and experience, and not simply follow a set algorithm. It was able to use the processes from prior studies, and either augment those algorithms with complementary procedures or employ entirely a new set of rules. The program was able to theoretically compensate for and process poorly-taken radiographs.[14,15]

However, at its current state of learning, the program has been 'taught' using only 400 images. This is below the 1,000-image minimum recommended by artificial neural network programmers for image identification. Generally, the more images used to teach the program, the better it becomes at image identification. In addition, the program was initially taught using high- resolution DICOM images. A self-diagnostic run at the end of the initial teaching phase showed an accuracy of 98.08%. This study used lower-quality digital camera images of analog radiographs, which may have affected the program's accuracy.

Hip fractures were selected as the initial target for the program since they are common, with a lifetime risk of 45–50% in women and 13–22% in men.[16] As previously mentioned, it is critical to diagnose and treat early to prevent the morbidities caused by prolonged immobilization.[17] We also reemphasize the gap in diagnosis and treatment where an estimated 10–30% of hip fractures are not detected on initial radiographs.[18] We were unable to find specific data to the setting of smaller primary and secondary health facilities, but we projected that the rates of missed hip fractures would be higher in areas where there is limited access to ancillary diagnostics, and the specialists to interpret the results. The program being tested aims to bridge this gap. At the program's current pilot version, the detection accuracy of 70% is comparable to the two prior programs targeting hip fractures. The high test sensitivity (97%) demonstrates the program's potential as a diagnostic screening tool. A limitation in the current study is the small sample size; thus, results of the computer interpretation were not analyzed according to hip fracture classifications (e.g., Garden, Pauwel). This means that the program sensitivity may still change if there are more non-displaced or minimally-displaced fractures to interpret. However, high test accuracy and sensitivity are not the main end points of a good diagnostic test, rather these serve as surrogates for relevant patient outcomes (e.g., functional capacity).[19] A good diagnostic test can influence patient outcomes. Aside from being highly sensitive, a reasonable test should be accessible and easy to use for it to be able to make an impact. Today's digital age offers exciting possibilities to answer these needs.

Over the past decade, imaging use persistently increased, with radiography as the initial diagnostic of choice in emergency department (ED) consultations. The high prevalence of missed hip fractures due to user-dependent error warrants a standard and objective procedure of interpreting radiographs. Given that the distribution of capable institutions with certified radiologists for musculoskeletal evaluation is centered on urban facilities, areas that are less equipped can benefit from AI technology. This technology will not replace the radiologist but would help augment the decision making and direction of radiographic examination coupled with telehealth. The National Telehealth Center of the Philippines already has an e-Health strategy that implements technology-based programs to improve access to quality health care. This includes the use of personal computers, tablets, and smart phones; with user-friendly interfaces linked to medical devices, which allow medical and allied medical health practitioners access to specialized health care via telemedicine.[4] Artificial Intelligence hip fracture detection is a step forward in this direction. It is compatible with the existing programs, and once fully developed with its ML capabilities, is poised to achieve the goal of bridging the gap towards accessible and affordable health care for all.

Future directions in program development include: 1) machine learning utilizing low-resolution images, which are the type of images that the target population of health care workers serving in the underserved areas can likely generate 2) variation of images to include non-displaced and minimally-displaced fractures, which are more likely to be missed by human interpreters 3) analysis of the program's ability to detect fractures categorized per different hip fracture classification 4) program's user interface and ease of use 5) migration and inter-operability between different program platforms.

Statement of Authorship

All authors participated in the data collection and analysis and approved the final version submitted.

Author Disclosure

All authors declared no conflicts of interest.

Funding Source

No funding support.

REFERENCES

1. Dhanwal DK, Dennison EM, Harvey NC, Cooper C. Epidemiology of hip fracture: worldwide geographic variation. Indian J Orthop. 2011 Jan; 45(1):15-22. doi: 10.4103/0019-5413.73656.
2. Carpintero P, Caeiro JR, Carpintero R, Morales A, Silva S, Mesa M. Complications of hip fractures: a review. World J Orthop. 2014 Sep 18; 5(4):402-11. doi: 10.5312/wjo.v5.i4.402.
3. Park YC, Um KS, Hong SP, Oh CW, Kim S, Yang KH. Preoperative computed tomography capsular sign for the detection of occult ipsilateral femoral neck fractures associated with femoral shaft fractures. Injury. 2020 Apr; 51(4):1051-6. doi: 10.1016/j.injury.2020.02.067.
4. Castellino RA. Computer aided detection (CAD): an overview. Cancer Imaging. 2005 Aug 23; 5(1):17-9. doi: 10.1102/1470-7330.2005.0018.
5. Eksi Z, Emre D, Cakiroglu M. Computer aided bone fracture detection. 2012 20th Signal Processing and Communications Applications Conference (SIU), 2012. doi:10.1109/siu.2012.6204644.
6. Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. N Engl J Med. 2016 Sep 29; 375(13):1216-9. doi: 10.1056/NEJMp1606181.

7. Cabitza F, Locoro A, Banfi G. Machine learning in orthopedics: a literature review. Front Bioeng Biotechnol. 2018 Jun 27; 6:75. doi: 10.3389/fbioe.2018.00075.

8. Bandyopadhyay O, Biswas A, Bhattacharya BB. Long-bone fracture detection in digital X-ray images based on digital-geometric techniques. Comput Methods Programs Biomed. 2016 Jan; 123:2-14. doi: 10.1016/j.cmpb.2015.09.013.

9. Custodio RG, Fernandez-Marcelo P, Magtubo KMP, Gavino A, Gonzales L, Gonzales A, et al. Adoption of a Telemedicine Device by Philippine Primary Care Health Workers. ResearchGate [Internet]. 2014 [cited 2017 Jan 30]. Available from: https://www.researchgate.net/publication/264708293_Adoption_of_a_Telemedicine_device_by_Philippine_Primary_Care_Health_Workers.

10. Viechtbauer W, Smits L, Kotz D, Budé L, Spigt M, Serroyen J, et al. A simple formula for the calculation of sample size in pilot studies. J Clin Epidemiol. 2015; 68(11):1375-9. doi: 10.1016/j.jclinepi.2015.04.014.

11. Agatonovic-Kustrin S, Beresford R. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J Pharm Biomed Anal. 2000 Jun; 22(5):717-27. doi: 10.1016/s0731-7085(99)00272-1.

12. Donnelley M, Knowles G. Computer aided long bone fracture detection., Proceedings of the Eighth International Symposium on Signal Processing and Its Applications. 2005; 1:175-8.

13. Donnelley M, Knowles G, Hearn T. A CAD System for Long-Bone Segmentation and Fracture Detection. In: Elmoataz A, Lezoray O, Nouboud F, Mammass D. (eds) Image and Signal Processing. ICISP 2008. Lecture Notes in Computer Science, vol 5099. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-69905-7_18

14. Zhou Y, Teomete U, Dandin O, Osman O, Dandinoglu T, Bagci U, et al. Computer-Aided Detection (CADx) for plastic deformation fractures in pediatric forearm. Comput Biol Med. 2016 Nov 1; 78: 120-5. doi: 10.1016/j.compbiomed.2016.09.013.

15. Nielsen M. Using neural nets to recognize handwritten digits [Internet]. Aug 2017 [cited April 2021]. Available from: http://neuralnetworksanddeeplearning.com/chap1.html

16. Court-Brown C, Heckman J, McQueen M, Ricci W, Tornetta P, McKee M. Rockwood and Green's Fractures in Adults. Philadelphia: Wolters Kluwer Health, 2015.

17. Rodriguez-Fernandez P, Adarraga-Cansino D, Carpintero P. Effects of delayed hip fracture surgery on mortality and morbidity in elderly patients. Clin Orthop Relat Res. 2011 Nov; 469(11):3218-21. doi: 10.1007/s11999-010-1756-z.

18. Gill SK, Smith J, Fox R, Chesser TJ. Investigation of occult hip fractures: the use of CT and MRI. S World J. 2013; 2013:830319. doi: 10.1155/2013/830319.

19. Cheng CT, Ho TY, Lee TY, Chang CC, Chou CC, Chen CC, et al. Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs. Eur Radiol. 2019 Oct; 29(10):5469-77. doi: 10.1007/s00330-019-06167-y.

APPENDIX

Appendix Table 1. AI Radiograph interpretation results

Image number	Left Hip Computer	Left Hip Actual	Interpretation	Right Hip Computer	Right Hip Actual	Interpretation
1	Normal	Normal	TN	Fractured	Fractured	TP
2	Fractured	Fractured	TP	Normal	Normal	TN
3	Fractured	Normal	FP	Fractured	Fractured	TP
4	Fractured	Fractured	TP	Fractured	Normal	FP
5	Fractured	Fractured	TP	Fractured	Normal	FP
6	Fractured	Fractured	TP	Fractured	Normal	FP
7	Fractured	Fractured	TP	Normal	Normal	TN
8	Fractured	Fractured	TP	Fractured	Normal	FP
9	Fractured	Normal	FP	Fractured	Fractured	TP
10	Fractured	Fractured	TP	Fractured	Normal	FP
11	Fractured	Normal	FP	Fractured	Fractured	TP
12	Fractured	Fractured	TP	Fractured	Normal	FP
13	Fractured	Fractured	TP	Normal	Normal	TN
14	Normal	Normal	TN	Fractured	Fractured	TP
15	Normal	Normal	TN	Fractured	Fractured	TP
16	Fractured	Normal	FP	Fractured	Fractured	TP
17	Fractured	Fractured	TP	Normal	Normal	TN
18	Fractured	Normal	FP	Fractured	Fractured	TP
19	Fractured	Fractured	TP	Fractured	Normal	FP
20	Fractured	Normal	FP	Fractured	Fractured	TP
21	Fractured	Fractured	TP	Normal	Normal	TN
22	Fractured	Normal	FP	Fractured	Fractured	TP
23	Normal	Fractured	FN	Normal	Normal	TN
24	Fractured	Fractured	TP	Fracture	Normal	FP
25	Fractured	Fractured	TP	Fractured	Normal	FP
26	Fractured	Normal	FP	Fractured	Fractured	TP
27	Normal	Normal	TN	Fractured	Fractured	TP
28	Normal	Normal	TN	Fractured	Fractured	TP
29	Fractured	Fractured	TP	Normal	Normal	TN
30	Normal	Normal	TN	Fractured	Fractured	TP
31	Normal	Normal	TN	Fractured	Fractured	TP
32	Normal	Normal	TN	Fractured	Fractured	TP
33	Fractured	Fractured	TP	Fractured	Normal	FP
34	Fractured	Fractured	TP	Fractured	Normal	FP
35	Fractured	Fractured	TP	Normal	Normal	TN
36	Fractured	Normal	FP	Fractured	Fractured	TP
37	Fractured	Fractured	TP	Fractured	Normal	FP
38	Normal	Normal	TN	Fractured	Fractured	TP
39	Fractured	Normal	FP	Fractured	Fractured	TP
40	Normal	Normal	TN	Fractured	Fractured	TP
41	Fractured	Fractured	TP	Fractured	Normal	FP
42	Fractured	Fractured	TP	Fractured	Normal	FP
43	Normal	Normal	TN	Fractured	Fractured	TP
44	Fractured	Normal	FP	Fractured	Fractured	TP
45	Fractured	Normal	FP	Fractured	Fractured	TP

TN – True Negative; TP – True Positive; FN – False Negative; FP – False Positive