

Mobile application as a complementary tool for differential diagnosis in Neuro-ophthalmology: A multicenter cross-sectional study

Pulikottil Wilson Vinny, Aastha Takkar¹, Vivek Lal¹, Madakasira Vasantha Padma^{2,3}, P N Sylaja⁵, Lakshmi Narasimhan⁶, Sada Nand Dwivedi⁴, Pradeep P Nair⁷, Thomas Iype⁸, Anu Gupta⁹, Venugopalan Y Vishnu²

Purpose: Drawing differential diagnoses to a Neuro-ophthalmology clinical scenario is a difficult task for a neurology trainee. The authors conducted a study to determine if a mobile application specialized in suggesting differential diagnoses from clinical scenarios can complement clinical reasoning of a neurologist in training. **Methods:** A cross-sectional multicenter study was conducted to compare the accuracy of neurology residents versus a mobile medical app (Neurology Dx) in drawing a comprehensive list of differential diagnoses from Neuro-ophthalmology clinical vignettes. The differentials generated by residents and the App were compared with the Gold standard differential diagnoses adjudicated by experts. The prespecified primary outcome was the proportion of correctly identified high likely gold standard differential diagnosis by residents and App. **Results:** Neurology residents ($n = 100$) attempted 1500 Neuro-ophthalmology clinical vignettes. Frequency of correctly identified high likely differential diagnosis by residents was 19.42% versus 53.71% by the App ($P < 0.0001$). The first listed differential diagnosis by the residents matched with that of the first differential diagnosis adjudicated by experts (gold standard differential diagnosis) with a frequency of 26.5% versus 28.3% by the App, whereas the combined output of residents and App scored a frequency of 41.2% in identifying the first gold standard differential correctly. The residents correctly identified the first three and first five gold standard differential diagnosis with a frequency of 17.83% and 19.2%, respectively, as against 22.26% and 30.39% ($P < 0.0001$) by the App. **Conclusion:** A ruled based app in Neuro-ophthalmology has the potential to complement a neurology resident in drawing a comprehensive list of differential diagnoses.

Key words: Clinical decision support system, differential diagnosis, medical education, neuro-ophthalmology

Neuro-ophthalmology poses a steep learning curve for residents in Neurology. Drawing differential diagnoses to a clinical scenario is a daunting task without the aid of clinical discussion involving teaching faculty, mentors, and literature references. Computer programs have shown to exhibit diagnostic accuracy which rivals that of expert clinicians.^[1,2] Differential diagnosis generators have been reported to complement clinical reasoning of a neurologist.^[3-5] A cross-sectional study was planned involving neurology residents, and a differential diagnoses generator mobile application (App) (Neurology Dx) to compare the accuracy in

drawing a comprehensive list of differential diagnoses from clinical vignettes involving Neuro-ophthalmology.

The authors (first and corresponding) created a differential diagnoses app (Neurology Dx) with an objective to compare the App's accuracy with that of Neurology residents, in drawing a comprehensive list of differential diagnoses in the context of a clinical vignette. We hypothesized that the App would be able to outperform and complement Neurology residents while drawing a comprehensive list of differential diagnoses to the same clinical vignettes.

Methods

Creation of app

Creation of app was achieved in a three-step process viz. (i) synthesizing Knowledge-base with data-points comprising of symptoms, signs, and neurological diagnosis (ii) defining the

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

Cite this article as: Vinny PW, Takkar A, Lal V, Padma MV, Sylaja PN, Narasimhan L, et al. Mobile application as a complementary tool for differential diagnosis in Neuro-ophthalmology: A multicenter cross-sectional study. Indian J Ophthalmol 2021;69:1491-7.

Access this article online

Website:

www.ijo.in

DOI:

10.4103/ijo.IJO_1929_20

Quick Response Code:



Neurology, Indian Naval Hospital Ship, Asvini, Mumbai, Maharashtra, ¹Neurology, Postgraduate Institute of Medical Education and Research, Chandigarh, ²Neurology, ³Neurosciences Centre and ⁴Biostatistics, All India Institute of Medical Sciences, New Delhi, ⁵Neurology, Sree Chitra Tirunal Institute of Medical Sciences and Technology, Thiruvananthapuram, Kerala, ⁶Neurology, Madras Medical College, Chennai, Tamil Nadu, ⁷Neurology, Jawaharlal Nehru Institute of Postgraduate Medical Education and Research, Puducherry, ⁸Neurology, Government Medical College Trivandrum, Kerala, ⁹Neurology, Govind Ballabh Pant Institute of Postgraduate Medical Education and Research, New Delhi, India

Correspondence to: Dr. Venugopalan Y. Vishnu, Assistant Professor (Neurology), Room No. 704, CN Centre, Seventh Floor, All India Institute of Medical Sciences, New Delhi, India. E-mail: vishnuvy16@yahoo.com

Received: 24-Jun-2020

Revision: 12-Jan-2021

Accepted: 23-Jan-2021

Published: 21-May-2021

associations between these data-points in the Knowledge-base and finally (iii) converting these data-points and their associations in knowledge-base into an App by the use of a Computer Language (Swift).

In the first step, an exhaustive search of published literature was undertaken to compile a list of data-points (Diagnoses, Symptoms, Signs, Imaging, and Lab findings) related to Neurology. The compiled list of data-points was then distilled to remove duplicate entries. In the next step, associations between these Data-Points in Knowledge-base were defined by assigning a frequency to pairs of Data Points by means of Fuzzy Logic. Fuzzy logic is a logical calculus that operates with many truth values. This is in contrast to classical logic that works with true and false values. This ability to handle multiple truth values for a given pair of patient attributes (Data-Points) gives fuzzy logic a unique advantage to factor in the uncertainties that are inherent to clinical scenarios. At the end of this exercise any single Data Point was associated with multiple other Data Points in the Knowledge-base. To illustrate an example, the Data Points representing diagnoses of Myasthenia Gravis and Botulism were associated with other Data Points in Knowledge-base viz. ptosis (Sign), double vision (Symptom), fatigability (Sign), pupillary changes (Sign), and flaccid weakness (Sign), with varying strengths of association. In turn, double vision (Symptom) was associated with Data Points representing multiple diagnoses namely Myasthenia Gravis, Tolosa-Hunt syndrome, PCOM aneurysm, etc., with differing strengths of association. In the third step, a computer algorithm was written by the first author in Computer Language (Swift version 3.0) to derive a list of DDx from any combination of user-selected Data Points (Symptoms, Signs, Lab and Imaging findings) available in the Knowledge-base. The algorithm harnesses the associations between Data Points in the Knowledge-base to derive DDx. The algorithm lists the derived DDx in a descending order of priority based on the strength of association between user-selected combination of Data Points and derived DDx [Fig. 1]. The algorithm's output was validated with previously published NEJM CPCs.^[6]

The knowledge-base comprising of the unique Data-Points and its associations, along with the computer coded algorithm were converted to a mobile App (Neurology Dx) and uploaded to AppStore.^[7] The final version (Neurology Dx version 2.1) was uploaded to AppStore on 13 June 2017. After commencement of study, no changes were made to the Knowledgebase or algorithm of the App (Neurology Dx) via further updates in AppStore, to prevent bias in the study and to ensure reproducibility of results.^[7] The detailed description of the creation of the App has been discussed in earlier papers on cognitive neurology, movement disorders, and stroke [Figs. 1 and 2].^[3-5] A multicenter cross-sectional study was designed involving seven teaching neurology centers in India. The critical elements of the study were creation of a differential diagnosis generator app, making of clinical vignettes in neuro-ophthalmology by subject experts, testing participating Neurology residents across the seven centres to draw a comprehensive list of differential diagnoses to these clinical vignettes, and generating a list of differential diagnoses by the App for the same clinical vignettes. Subject experts derived a list of gold standard differentials to clinical vignettes from their clinical experiences. The differential diagnoses derived by the residents and App were then compared with the gold

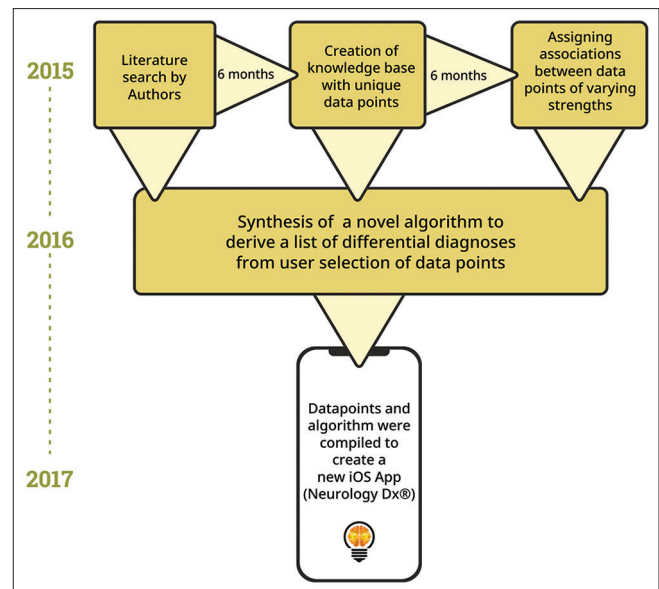


Figure 1: Pathway depicting the creation of the App

standard differentials synthesized by the subject experts. Two subject experts were involved for Neuro-ophthalmology (one Assistant Professor and Professor of Neurology who specialize in Neuro-ophthalmology). The protocol of the study was prospectively registered in the Clinical Trial Registry of India (CTRI/2017/06/008838).

A panel of subject experts created 60 clinical vignettes in Neuro-ophthalmology [Supplemental File]. The clinical vignettes were so designed to elicit broad differential diagnoses akin to the scenario of usual clinical practice. These clinical vignettes were graded by experts to be of varying grades of difficulty viz mild, moderate, and severe. These vignettes were randomly grouped into sets of four. A flow diagram depicting the creation of the App and preparation of the vignette is shown in Fig. 3. The study was conducted from July 15 to August 15, 2017, in seven tertiary care teaching neurology centers in India [Fig. 4]. These centers were AIIMS New Delhi, PGIMER Chandigarh, GIPMER Delhi, JIPMER Puducherry, MMC Chennai, and GMC Trivandrum. We invited all the neurology residents from participating centers to enroll in the study. All participating residents possessed a recognized doctoral degree in Internal Medicine from an institute recognized by the Medical Council of India and were pursuing an ongoing residency program in Neurology at the participating centers. Fellows and faculty in Neurology were excluded from the study. Formal consent was obtained. The participants were not allowed to use mobile phones, computers or the internet while attempting to draw a list of differential diagnoses from the clinical vignettes.

All residents were instructed to write down all the relevant differential diagnosis to a clinical vignette on a sheet of paper in decreasing order of priority. A time limit to attempt clinical vignettes was not fixed. Four sets (a total of 60 clinical vignettes) were used in the study. Each resident attempted one set comprising of 15 clinical vignettes of varying difficulty (mild—5, moderate—5, severe—5). The differential diagnosis listed by the residents for each vignette was recorded in duplicate. Each clinical vignette was also fed into the App

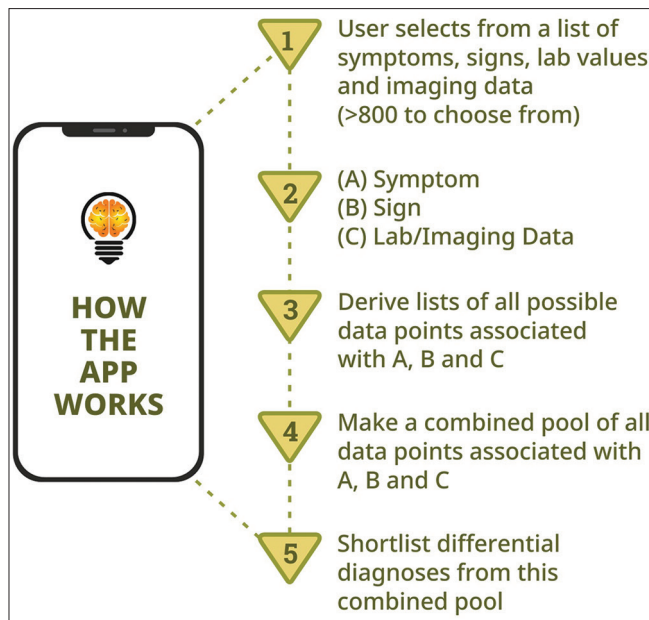


Figure 2: App's working algorithm

to derive App suggested differentials, which was tabulated separately. All the differentials given by the residents and App for a particular clinical vignette were pooled and shown to experts in Neuro-ophthalmology. The chosen experts were full-time faculty in Neurology who specialize in Neuro-ophthalmology. Each expert had a minimum of 10 years' experience in training Neurology residents. The experts grouped the differentials into "high likely", "moderate likely", "less likely", and "wrong answers". While deriving the gold standard differential diagnosis, the experts were blinded to the source of pooled differentials presented to them. Experts were also at liberty to add new differentials not available in the pooled list presented to them. Differences of opinion between experts were resolved by mutual consultation. The list of differentials given by the residents and suggested by the App were separately compared with the gold standard differentials derived by the experts.

The prespecified primary outcome was the proportion of correctly identified high likely gold standard differential diagnosed by residents and App. The secondary outcomes included whether the residents and App could accurately identify the first differential of the gold standard differential as their first differential. Besides, other secondary outcomes included proportions of correctly identified differential from various subsets of gold standard differentials viz "high likely differentials", "first three high likely differential", "First five high likely differentials" and "combined list of high and moderate likely differentials" by residents and App, respectively. The subgroup analysis conducted included diagnostic accuracy based on year of residency (prespecified) and level of difficulty of clinical vignettes (*post hoc*). The sample size of 1500 neuro-ophthalmology clinical vignettes was estimated based on a previous study comparing the diagnostic accuracy of physicians and a computer algorithm.^[8] We compared the proportion of correctly identified differentials between the App and neurology residents using proportion test and documented the absolute difference with 95% confidence

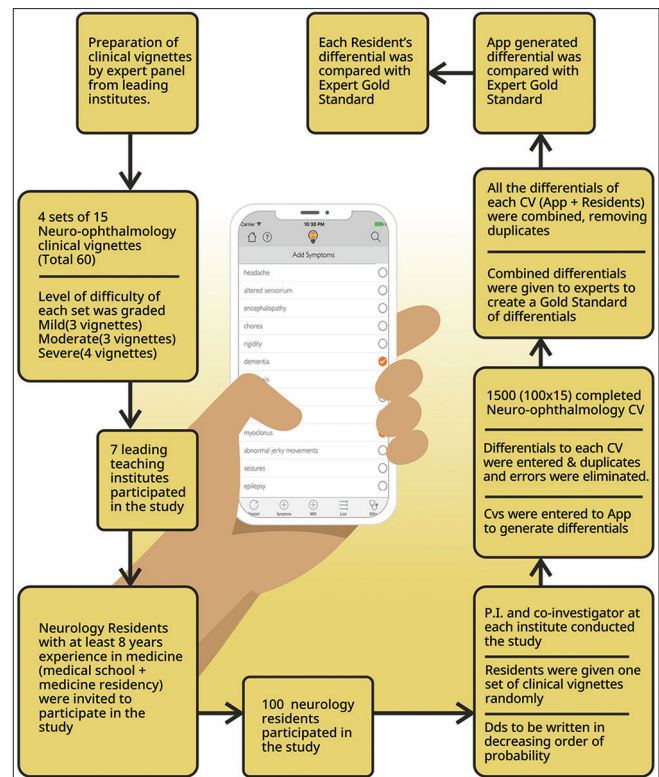


Figure 3: Overall study workflow

interval. The same method was used for comparison of primary and secondary outcomes.

Results

We identified 122 Neurology residents who qualified for the inclusion and exclusion criteria. Of these, seven residents were employed on emergency duty at the time of conducting the study and 15 residents were out of station on leave. The remaining 100 residents were enrolled in the study after obtaining informed consent. At the end of the study, a total of 1500 clinical vignettes in Neuro-ophthalmology had been attempted by 100 participating residents (15 vignettes X 100 residents). Of the 100 residents who participated in the study, 45 were in the first year of residency, 33 in the second year, and 22 in the third year of residency [Fig. 4]. When high likely gold standard differentials were compared with that of the differentials given by residents and App, the frequency of correctly identified high likely differentials by residents was 19.42% versus 53.71% by the App ($P < 0.0001$). In a subgroup analysis, frequency of correct identification of high likely gold standard differentials amongst the first year, second-year, and third-year residents was done. The first-, second-, and third-year residents scored a frequency of 16.31%, 20.01%, and 20.8%, respectively, as against 53.71% by the App, $P < 0.0001$. [Tables 1 and 2].

When high and moderate likely gold differentials were combined to detect the accuracy of diagnosis, residents and App correctly diagnosed with a frequency of 12.21% and 41.93%, respectively ($P < 0.0001$). Subgroup analysis for correct identification of high likely and moderate likely gold standard differentials showed that the first year, second year and

third-year residents scored a frequency of 10.16%, 12.19%, and 16.3%, respectively, as against 41.93% by the App ($P < 0.0001$). Residents identified the first gold standard differential as their first differential with a frequency of 26.5% versus 28.3% by the App. We analyzed the results where the output of the App and residents were combined (first differential given by residents and suggested by the App were taken together) and compared against the first gold standard differential. The combined output of residents and App correctly identified the first gold standard differential with a frequency of 41.2%.

The residents correctly identified the first three and first five gold standard differentials with a frequency of 17.83% and 19.2%, respectively, as against 22.26% and 30.39% ($P < .0001$) by the App. A post hoc subgroup analysis of differential diagnoses generated on Neuro-ophthalmology clinical vignettes based on the level of difficulty of vignettes was performed [Table 3]. When only mild grade clinical vignettes were considered, the App correctly identified high likely gold standard differentials with a frequency of 47.87% versus 29.44% by the residents ($P = 0.0001$). Similarly, for mild grade clinical vignettes the App correctly identified “combined high and moderate likely gold standard differentials”, “first gold standard differential”, “first three gold standard differentials” and “first five gold standard differentials with a frequency of 38.71%, 55%, 22.34%, and 28.72%, respectively, as against 18.94% ($P < 0.0001$), 40.8%, 25.52%, and 28.89%, respectively, by the residents.

When only moderate grade clinical vignettes were considered, the App correctly identified high likely gold standard differentials with a frequency of 51.02% versus 18.52% by the residents ($P < 0.0001$). Similarly, for moderate grade clinical vignettes the App correctly identified “combined high and moderate likely gold standard differentials”, “first gold standard differential”, “first three gold standard differentials” and “first five gold standard differentials with a frequency of 41.56%, 25%, 22.45%, and 27.55%, respectively, as against 11.12% ($P < 0.0001$), 27.6%, 17.73%, and 18.4%, respectively, by the residents. When only severe grade clinical vignettes were considered, the App correctly identified high likely gold standard differentials with a frequency of 61.54% versus 10.46% by the residents ($P < 0.0001$). Similarly, for severe grade clinical vignettes the App correctly identified “combined high and moderate likely gold standard differentials”, “first gold standard differential”, “first three gold standard differentials” and “first five gold standard differentials with a frequency of 45.18%, 5%, 21.99%, and 35.17%, respectively, as against 6.33% ($P < 0.0001$), 11%, 10.33% and 10.46% ($P < 0.0001$), respectively, by the residents.

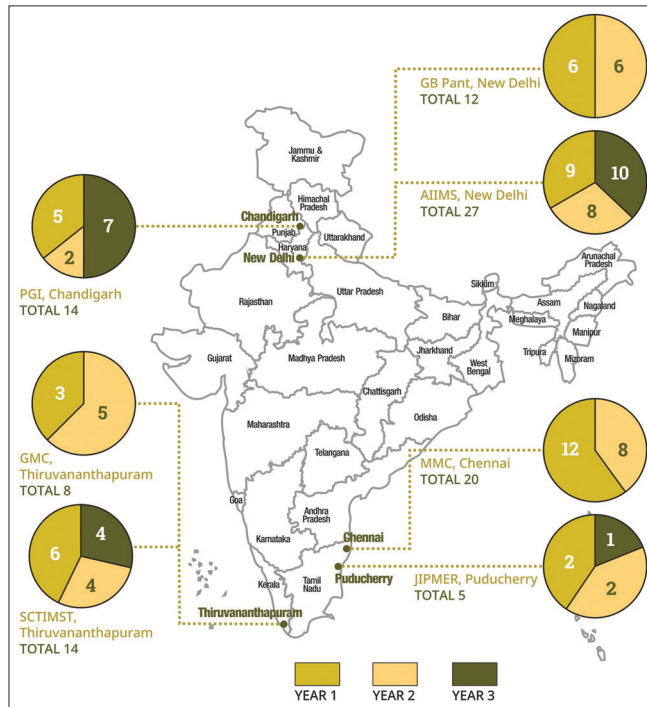


Figure 4: Participating centers with number of residents in each year of residency (shades) and the set of clinical vignettes (CV series I-IV) used in each center (numbers)

Table 1: Differential diagnosis in Neuro-ophthalmology clinical vignettes: Neurology Residents vs. App (Neurology Dx®)

Neurology Residents (100)	Residents	Neurology Dx	Absolute difference (95% CI)	P
High Likely Differentials (HLD) (Correctly identified)	1408/7251 (19.42%)	152/283 (53.71%)	34.29 (28.4-40.1)	<.0001
High and Moderate Likely Differentials-Combined	2084/17074 (12.21%)	278/663 (41.93%)	29.72 (26.0-33.5)	<.0001
High Likely Differential: 1 st Differential of Gold standard is correctly identified as first differential	397/1500 (26.5%)	17/60 (28.3%)	1.83 (-8.3-14.5)	=0.75
High Likely Differential: 1 st Differential of Gold standard is correctly identified as first differential (Combined: App + resident)	618/1500 (41.2%)			
No. of First 3 differentials identified from total gold standard HLD	1293/7251* (17.83%)	63/283* (22.26%)	4.43 (-0.12-9.70)	=0.057
No. of First 3 differentials identified from maximum possible gold standard HLD (number of entries x3)	1293/4500** (28.73%)	63/180** (35%)	6.27 (-0.45-13.60)	=0.069
No. of First 5 differentials identified from total gold standard HLD***	1392/7251 (19.2%)	86/283 (30.39%)	11.19 (6.04-16.85)	<.0001

*Denominator includes all the high likely gold standard answers used in the evaluation of residents (1500 clinical vignettes- 15 x 100) and NeurologyDx (60 clinical vignettes-15 x 4). ** Denominator includes maximum possible high likely gold standard answers in the first three differentials (1500 x 3=4500 for residents and 60 x 3=180 for NeurologyDx). ***Denominator of total high likely gold standard answers in the first five differentials is less than the maximum possible differentials (1500 x 5=7500 for residents and 60 x 5=300 for App). Hence maximum possible differentials are not used in the denominator

Table 2: Prespecified Subgroup Analysis of differential diagnoses generated on Neuro-ophthalmology clinical vignettes based on the year of neurology residency

First year Resident (45)	Residents	Neurology Dx	Absolute difference (95% CI)	P
High Likely Differentials (HLD) (Correctly identified)	538/3299 (16.31%)	152/283 (53.71%)	37.4 (31.4-43.3)	<.0001
High and Moderate Likely Differentials-Combined	780/7677 (10.16%)	278/663 (41.93%)	31.77 (28.0-35.6)	<.0001
High Likely Differentials 1 st Differential of Gold standard is correctly identified as first Differential	172/675 (25.48%)	17/60 (28.3%)	2.8 (-7.6-15.6)	=0.632
No. of First 3 differentials identified from total gold standard HLD	502/3299* (15.22%)	63/283* (22.26%)	7.04 (2.4-12.4)	=0.00018
No. of First 3 differentials identified from maximum possible gold standard HLD (number of entries x 3 ; 675 x 3=2025)	502/2025** (24.79%)	63/180** (35%)	10.21 (3.4-17.7)	=0.0026
No. of First 5 differentials identified from gold standard HLD***	534/3299 (16.19%)	86/283 (30.39%)	14.20 (9.0-20.0)	<.0001
Second year Resident (33)				
High Likely Differentials (HLD) (Correctly identified)	475/2374 (20.01%)	152/283 (53.71%)	33.7 (27.7-39.6)	<.0001
High and Moderate Likely Differentials-Combined	679/5569 (12.19%)	278/663 (41.93%)	29.7 (25.9-33.6)	<.0001
High Likely Differentials 1 st Differential of Gold standard is correctly identified as first differential	121/495 (24.44%)	17/60 (28.3%)	3.86 (-6.7-16.8)	=0.514
No. of First 3 differentials identified from gold standard HLD	440/2374* (18.53%)	63/283* (22.26%)	3.73 (-1.0-9.2)	=0.130
No. of First 3 differentials identified from maximum possible gold standard HLD (number of entries x 3 ; 495 x 3=1485)	440/1485** (29.63%)	63/180** (35%)	5.37 (-1.6-12.9)	0.139
No. of First 5 differentials identified from gold standard HLD***	470/2374 (19.80%)	86/283 (30.39%)	10.59 (5.3-16.4)	<.0001
Third year Resident (22)				
High Likely Differentials (HLD) (Correctly identified)	298/1432 (20.8%)	152/283 (53.71%)	32.91 (26.7-39.0)	<.0001
High and Moderate Likely Differentials-Combined	624/3828 (16.3%)	278/663 (41.93%)	25.6 (21.7-29.6)	<.0001
High Likely Differentials 1 st Differential of Gold standard is correctly identified as first answer	104/330 (31.52%)	17/60 (28.3%)	3.22 (-10.1-14.3)	=0.620
No. of First 3 differentials identified from gold standard HLD	350/1578* (22.18%)	63/283* (22.26%)	0.08 (-4.9-5.6)	=0.976
No. of First 3 differentials identified from maximum possible gold standard HLD (number of entries x 3 ; 330 x 3=990)	350/990** (33.33%)	63/180** (35%)	1.67 (-5.57-9.43)	=0.663
No. of First 5 differentials identified from gold standard HLD***	387/1578 (24.52%)	86/283 (30.39%)	5.87 (0.35-11.8)	=0.037

*Denominator includes all the high likely gold standard answers used in the evaluation of residents. **Denominator includes maximum possible high likely gold standard answers in the first three differentials. ***Denominator of total high likely gold standard answers in the first five differentials is less than the maximum possible differentials. Hence maximum possible differentials are not used in the denominator

Discussion

Neurology residents often find it challenging to reach a clinical diagnosis. The learning process involves two important skills: the skill to identify clinical signs (includes fundus examination) correctly and ability to integrate the elicited symptoms and signs with a previously learned knowledge base for making a comprehensive list of differential diagnoses. In a study comparing physicians with differential diagnosis algorithm in internal medicine, physicians vastly outperformed computer algorithms in diagnostic accuracy (84.3% vs 51.2% correct diagnosis in the top 3 listed differentials). Physicians in the study also gave incorrect diagnosis in about 15% of cases.^[8] In a systematic review and meta-analysis to investigate the efficacy and utility of DDX generators, DDX generators did not demonstrate improved diagnostic retrieval compared to clinicians.^[9] However small improvements were seen in the before and after studies where clinicians had the opportunity to revisit their diagnoses following DDX generator consultation.^[9]

In the present study, we have tried to compare the diagnostic accuracy of an app created de-novo (Neurology Dx) with that of Neurology Residents, as regards their ability to generate a comprehensive list of differential diagnoses given the same set of clinical vignettes. Neuro-ophthalmology experts separately derived the appropriate differentials to all clinical vignettes that served as the gold standard. As the study was conducted in a staggered manner, the sets of vignettes were used randomly in the participating centers. We achieved a sample size of 1500 completed clinical vignettes (15 vignettes X100 residents) in the study.

Neurology Dx identified 54% of high likely gold standard differentials correctly compared to 20% by the residents. The experts who constructed the clinical vignettes consciously made them open-ended so that more differentials are generated. This is representative of usual clinical practice where the funnel of differential diagnoses is broad, to begin with, and narrows down with successive clinical findings and lab investigations. The limitation of this approach was low diagnostic accuracy

Table 3: Post hoc subgroup analysis of differential diagnoses generated on Neuro-ophthalmology clinical vignettes based on the level of difficulty of vignettes

Mild Grade	Residents	Neurology Dx	Absolute difference (95% CI)	P
High Likely Differentials (HLD) (Correctly identified)	691/2347 (29.44%)	46/94 (47.87%)	18.43 (8.4-28.6)	=0.0001
High and Moderate Likely Differentials-Combined	1049/5540 (18.94%)	84/217 (38.71%)	19.77 (13.5-26.5)	<.0001
High Likely Differentials 1 st Differential of Gold standard is correctly identified as first Differential	204/500 (40.8%)	11/20 (55%)	14.2 (-7.0-33.8)	=0.207
No. of First 3 differentials identified from total gold standard HLD	599/2347* (25.52%)	21/94* (22.34%)	3.18 (-6.4-10.6)	=0.487
No. of First 3 differentials identified from maximum possible gold standard HLD (number of entries x 3 ; 500 x 3=1500)	599/1500** (39.93%)	21/60** (35%)	4.93 (-7.9-16.1)	=0.444
No. of First 5 differentials identified from gold standard HLD***	678/2347 (28.89%)	27/94 (28.72%)	0.17 (-9.8-8.6)	=0.972
Moderate grade				
High Likely Differentials (HLD) (Correctly identified)	469/2533 (18.52%)	50/98 (51.02%)	32.5 (22.6-42.3)	<.0001
High and Moderate Likely Differentials-Combined	708/6368 (11.12%)	101/243 (41.56%)	30.44 (24.4-36.8)	<.0001
High Likely Differentials 1 st Differential of Gold standard is correctly identified as first differential	138/500 (27.6%)	5/20 (25%)	2.6 (-19.6-17.0)	=0.799
No. of First 3 differentials identified from gold standard HLD	449/2533* (17.73%)	22/98* (22.45%)	4.72 (-2.6-14.0)	=0.232
No. of First 3 differentials identified from maximum possible gold standard HLD (number of entries x 3; 500 x 3=1500)	449/1500** (29.93%)	22/60** (36.67%)	6.74 (-4.6-19.6)	=0.265
No. of First 5 differentials identified from gold standard HLD***	466/2533 (18.4%)	27/98 (27.55%)	9.15 (1.1-18.8)	=0.023
Severe Grade				
High Likely Differentials (HLD) (Correctly identified)	248/2371 (10.46%)	56/91 (61.54%)	51.08 (40.7-60.5)	<.0001
High and Moderate Likely Differentials-Combined	327/5166 (6.33%)	93/203 (45.81%)	39.48 (32.7-46.4)	<.0001
High Likely Differentials 1 st Differential of Gold standard is correctly identified as first answer	55/500 (11%)	1/20 (5%)	6 (-12.8-11.1)	=0.397
No. of First 3 differentials identified from gold standard HLD	245/2371* (10.33%)	20/91* (21.99%)	11.66 (4.3-21.3)	=0.0004
No. of First 3 differentials identified from maximum possible gold standard HLD (number of entries x 3; 500 x 3=1500)	245/1500** (16.33%)	20/60** (33.33%)	17 (6.2-29.7)	=0.0006
No. of First 5 differentials identified from gold standard HLD***	248/2371 (10.46%)	32/91 (35.17%)	24.71(15.6-35.0)	<0.0001

*Denominator includes all the high likely gold standard answers used in the evaluation of residents. **Denominator includes maximum possible high likely gold standard answers in the first three differentials. ***Denominator of total high likely gold standard answers in the first five differentials is less than the maximum possible differentials. Hence maximum possible differentials are not used in the denominator

amongst the residents, when seen as a proportion of fewer generated differentials given by residents against a large number of gold standard differentials. When we considered only the first correct high likely expert answer, there was no difference between the residents and the App. If the clinical vignettes were designed in such a way as to generate only 3–5 differentials, the overall diagnostic accuracy would have been much higher. Majority of residents were first-year neurology residents, and hence their knowledge base in Neuro-ophthalmology may not be adequate. The performance of residents improved with the increasing number of years in residency. The app performed better when the severity of clinical vignettes increased. When we combined the differentials given by the App and residents,

the overall diagnostic accuracy showed a marked improvement compared to their performance metrics individually. The App may thus serve as a complementary tool in the hands of neuro-ophthalmology residents in training. Further, similar apps based on rules framed by experts to derive differential diagnosis in different fields of Neurology is a need of the hour in resource-limited and busy clinical practice settings.

The App has built-in features whereby the App may learn to perform better when trained with the gold standard differentials. Similarly, the knowledge base can be improved for better accuracy with each successive versions of the App. These features will significantly boost the performance of the App. In our study, we did not use these features as it was difficult to

study them over the prespecified study duration. We used the untrained and unmodifiable version of the App downloaded from Appstore. Future studies wherein the App trained by way of clinical vignettes when compared with residents and specialists may reveal exciting results.

Conclusion

A ruled based app in Neuro-ophthalmology has the potential to complement a neurology resident in drawing a comprehensive list of differential diagnoses. Our study shows the potential of such bedside tools in flattening the steep learning curve in Neuro-ophthalmology. A differential diagnoses generator based on an expert algorithm designed by academic neurologists will find immense use in resource-limited settings, busy clinical practice, and free up precious time for the clinician to engage in patient care.

Acknowledgements

Ms. Abhilasha Krishnamurthy (December Design Studio) for providing illustrative diagrams.

Ethical approval

The study was approved by AIIMS Institute Ethics Committee.

Financial support and sponsorship

Nil.

Conflicts of interest

Dr. VPW: Copyright owner of App is spouse of VPW

References

1. Berner ES, Webster GD, Shugerman AA, Jackson JR, Algina J, Baker AL, *et al.* Performance of four computer-based diagnostic systems. *N Engl J Med* 1994;330:1792-6.
2. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. *JAMA* 2018;320:2199-200.
3. Vinny PW, Gupta A, Modi M, Srivastava MVP, Lal V, Sylaja PN, *et al.* Head to head comparison between neurology residents and a mobile medical application (App) for diagnostic accuracy in cognitive neurology. *QJM* 2019;112:591-8.
4. Vinny PW, Rajan R, Goyal V, Padma MV, Lal V, Sylaja PN, *et al.* Deducing differential diagnoses in movement disorders: Neurology residents versus a novel mobile medical application (Neurology Dx). *Ann Mov Disord* 2019;2:115-25.
5. Vinny PW, Padma MV, Sylaja PN, Kesav P, Lal V, Narasimhan L, *et al.* Comparison of diagnostic accuracy in vascular neurology between neurology residents and a neurology differential diagnosis app: A multi-center cross-sectional observational study. *J Stroke Med* 2020;3:21-7.
6. Vishnu VY, Vinny PW, Balaini N, Modi M, Radotra BR, Lal V. Computer algorithm as a supporting tool for clinical decision making on a mobile platform. *J Neurol Sci* 381:S751-2.
7. Available from: <https://itunes.apple.com/us/app/neurology-dx/id1224462861?mt=8>.
8. Semigran HL, Levine DM, Nundy S, Mehrotra A. Comparison of physician and computer diagnostic accuracy. *JAMA Intern Med* 2016;176:1860-1.
9. Riches N, Panagioti M, Alam R, Cheraghi-Sohi S, Campbell S, Esmail A, *et al.* The effectiveness of electronic differential diagnoses (DDX) generators: A systematic review and meta-analysis. *PLoS One* 2016;11:e0148991.