ORIGINAL PAPER

Assessing ASPECTS and ICH score reliability on NCCT scans via SMART INDIA App and PACS by neurologists and neuro-radiologists

Biswamohan Mishra[1] · Ayush Agarwal[1] · Ajay Garg[2] · Yamini Antil[1] · Sakshi Sharma[1] · Aprajita Parial[1] · Nilima Nilima[3] · Venugopalan Y. Vishnu[1] · MVPadma Srivastava[1]

Abstract

Purpose Inter-rater reliability is a critical aspect of stroke image interpretation. This study aims to investigate inter-rater reliability between neurologists and neuro-radiologist when assessing Alberta Stroke Program Early CT Score (ASPECTS) and Intracerebral Haemorrhage (ICH) scores using a mobile application (SMART INDIA App – by neurologists) and the Picture Archiving and Communication System (PACS – by neuro-radiologist).

Methods Adult patients diagnosed with ischemic or haemorrhagic stroke were included in this study. Two Neurologists (R1 and R2) assessed the ASPECTS and ICH scores by viewing the SMART INDIA App. A neuroradiologist expert (R3) assessed the same using PACS. Kappa statistics are presented for agreement between the Raters.

Results 100 consecutive patients each of Acute Ischemic stroke (AIS) and ICH were included. A significant agreement in the total ASPECTS between the Raters 1 and 2 (0.85(95% CI—0.775,0.926)), Raters 1 and 3(0.76(95% CI—0.671,0.857)) and Raters 2 and 3(0.76(95% CI—0.673,0.857)) was noted. A good agreement was ascertained between Rater 1 and 3, even though the devices were different. A similar excellent agreement was also noted in assessing the ICH score. The study's findings indicate that neurologists using different devices and platforms demonstrated good to excellent agreement with the neuro-radiologist (considered the gold standard) when estimating ASPECTS and ICH scores.

Conclusion The interpretations of neurologists using the SMART INDIA App can be deemed valid and reliable from the neuroimaging viewpoint. These results contribute to our understanding of the feasibility and reliability of app-based image evaluation in the assessment of stroke.

Keywords Inter-rater reliability · ASPECTS · ICH score · Smartphone app

1 Introduction

Inter-rater reliability is a measure of how well two or more raters agree on the same set of data [1, 2]. Inter-rater reliability assessment is crucial in the context of stroke imaging to ensure consistent and accurate interpretation of imaging findings [3]. The ASPECTS (Alberta Stroke Program Early CT Score) is a 10-point semiquantitative topographic scoring system used to assess early ischemic changes on non-contrast computed tomography (NCCT) scans in patients with anterior circulation hyperacute ischemic stroke [4]. Initially designed to identify those patients most likely to experience the greatest clinical benefit from thrombolytic therapy [5], ASPECTS is now a widely recommended crucial selection criteria for imaging selection for endovascular therapies [6, 7, 8].

The ICH (intracerebral haemorrhage) score is a simple and reliable grading scale that helps stratify the risk and predict 30-day mortality among patients with intracerebral haemorrhage [9]. The score is based on five clinical features: Glasgow Coma Scale (GCS) score, age, infratentorial origin of haemorrhage, volume of haemorrhage, and presence of intraventricular haemorrhage. The ICH score ranges from 0 to 6, with higher scores indicating a worse prognosis [10].

✉ Venugopalan Y. Vishnu
vishnuvy16@yahoo.com

1. Department of Neurology, All India Institute of Medical Sciences, New Delhi, India

2. Department of Neuroimaging and Interventional Neuroradiology, All India Institute of Medical Sciences, New Delhi, India

3. Department of Biostatistics, All India Institute of Medical Sciences, New Delhi, India

Reliable and consistent assessment of ASPECTS and ICH scores on NCCT brain scans is essential for making accurate clinical decisions, determining treatment strategies, and ensuring the validity of research studies. Robust inter-rater reliability assessment strengthens confidence in the interpretations and facilitates the comparability of findings across different studies and healthcare settings [3].

SMART INDIA is a multi-centre, open-label cluster-randomized trial that aims to enhance acute stroke care delivery in resource-limited settings in rural India [11]. The "SMART-India App," a mobile application, was developed with the primary goal of providing cost-effective Telestroke services, connecting neurologists and physiotherapists with physicians in district hospitals. A comprehensive pilot testing phase was conducted across all nodal centers to ensure the efficacy of the application.

The Smart India app was designed and developed by onsite developers engaged in the research project over an 8-month period from December 2020 to July 2021 in New Delhi, India. The android version of the SMART-INDIA Stroke-App is available on the Play Store, with access restricted to study participants, including physicians, neurologists, physiotherapists, and data entry operators (DEO).

Registration on the Smart India app involves neurologists, physicians, physiotherapists, research coordinators, and DEOs from recruited nodal centers and district hospitals. The nodal centers are interconnected with their respective district hospitals, facilitating seamless communication. When a patient with acute stroke arrives at an activated center, the physician at the district hospital immediately logs into the SMART INDIA APP and begins entering the patient's details. Neurologists receive instant notification alerts through the app, enabling real-time interaction between neurologists and physicians via the chat box. After planning acute stroke management, the neurologist concludes the chat, and a call request for telerehabilitation is activated, connecting to a study/notified physiotherapist. This integrated approach enhances timely communication and decision-making for the benefit of stroke patients. Assessing the inter-rater reliability between neurologists and neuro-radiologists is a crucial step before implementing the SMART India App for widespread use.

The purpose of this study is to evaluate the inter-rater reliability between neurologists' assessment of CT images uploaded to the SMART INDIA App and the neuroradiologist's assessment of the same CT images via PACS, for both the ASPECTS and ICH score, in the patients visiting a tertiary care centre. We hypothesize that there will be an acceptable level of inter-rater reliability between the neurologists and neuro-radiologists in this context.

2 Methodology

2.1 Study design and population

This study is a prospective observational study, performed in accordance with STROBE Guidelines. The study was conducted at All India Institute of Medical Sciences (AIIMS), a tertiary care centre in north India, between October 2022 and January 2023. Ethics approval was obtained from AIIMS Institutional Ethics Committee vide IEC-74/07.02.2020. Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research. The inclusion criteria for the study were adult patients diagnosed with ischemic or haemorrhagic stroke and with NCCT brain available in the institutional Picture Archiving and Communication System (PACS). Patients with stroke from other causes such as cerebral venous thrombosis, trauma, and posterior reversible encephalopathy syndrome (PRES) were excluded. Informed written consent was obtained from all patients or their caregivers. Patients were recruited from the outpatient services or emergency department of this tertiary care centre.

2.2 Data collection and analysis

A schematic diagram of the workflow employed in this study is shown in Fig. 1. Two investigators (YA and VYV) dichotomized the patients into ischemic stroke and hemorrhagic stroke based on their clinical presentation and imaging findings. The baseline characteristics of the included patients were recorded using Microsoft Excel version 16.73. To reduce variation in the image quality, NCCT scans of all the included patients were recorded using one device (SAMSUNG Galaxy Tab A; 8MP camera; Full HD (1080p) Recording) from a single monitor (Generic PnP Monitor on Standard VGA Graphics Adapter, Resolution – 1366×768) and by a single investigator YA at a rate of 1 frame per second, from the PACS. The recorded files were uploaded to the SMART INDIA App. The Unique IDs of the patients were shared with two neurologists (BM—Rater 1 (R1) and AA—Rater 2) and a neuroradiologist expert AG (>25 years of experience, GOLD Rater—Rater 3). The neurologists (R1 and R2) with a combined >10 years of experience, independently assessed the ASPECTS score and ICH score using the uploaded videos in the SMART INDIA App. The neuroradiologist expert (R3) assessed the scores using the original scan images in the PACS at a departmental workstation of the hospital. Each Rater recorded their assessments in Excel under suitable headings for the respective scores (Fig. 1).

The assessments from the individual Raters were collected by the investigators YA and VYV. The findings of
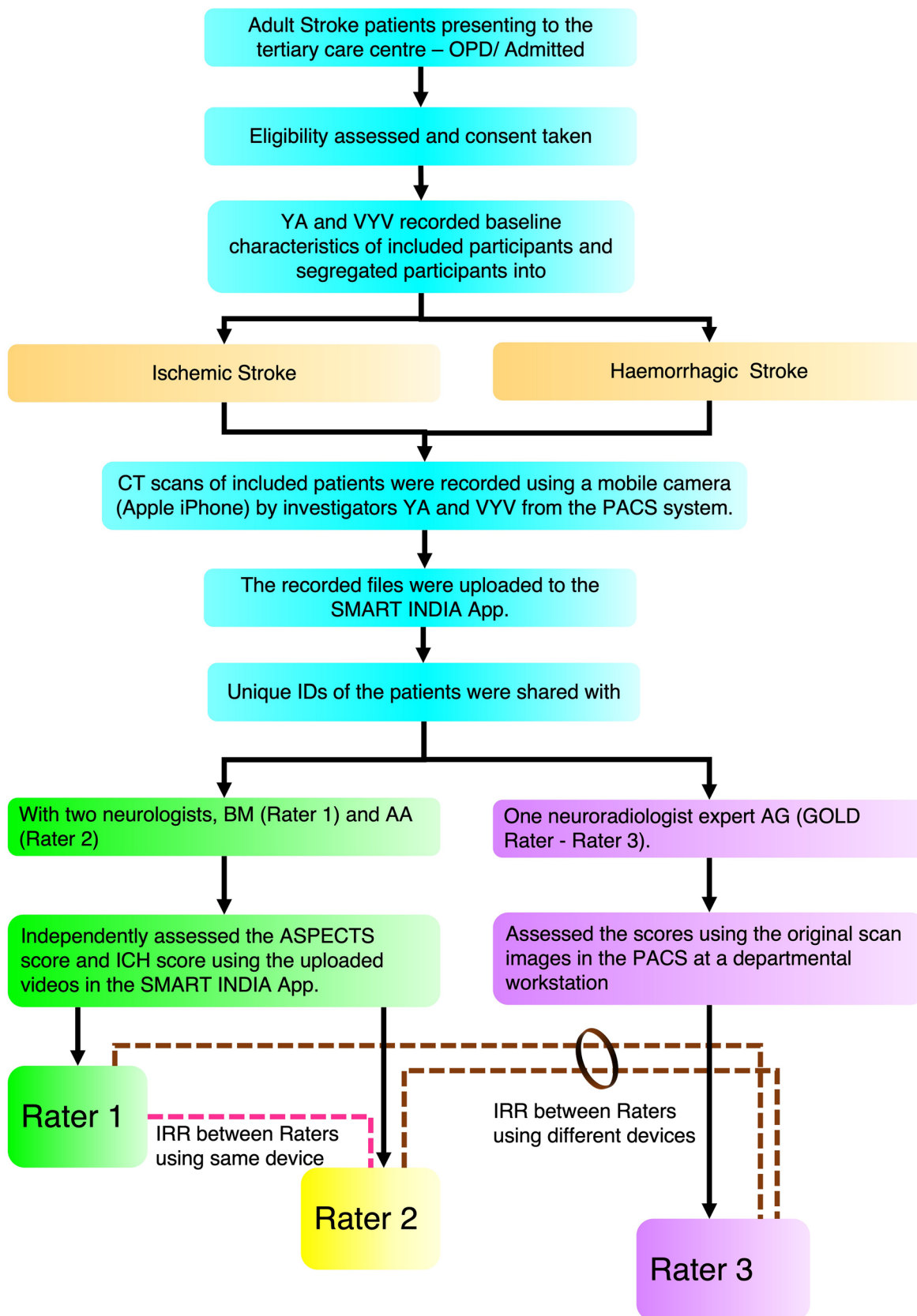
Fig. 1 shows the workflow employed in this study

the individual Raters were shared with a statistics expert (NN) for comparison and statistical analysis of the ratings. The identity of the Raters was kept blinded to maintain objectivity and avoid bias in the analysis.

2.3 Outcomes

Primary Outcome: To assess the inter-rater reliability (IRR) between the neurologists and the neuro-radiologists in the total ASPECTS and ICH score.

Secondary Outcome: To assess the IRR among the neurologists and the neuroradiologist on the various sub-components of ASPECTS and ICH score and the neuro-radiologist on the various sub-components of ASPECTS and ICH score.

3 Statistical analysis

The data is presented either as mean ± standard deviation or median and interquartile range (IQR). Weighted kappa statistic (95% confidence interval) are presented for inter-observer agreement for total scores and a Cohen's κ was calculated for individual regions and post dichotomisation of ASPECT score. Stata version 16 (StataCorp, College Station, Texas, USA) was used for analysis. Bootstrap kappa with 1000 replications was used to obtain the 95% confidence interval for the kappa statistic.

3.1 Baseline characteristics

100 consecutive patients with Acute Ischemic stroke (AIS) and 100 consecutive patients with haemorrhagic stroke (ICH – Intra-cerebral haemorrhage) were included in this study. The average age of the patients with ischaemic stroke was 61.81 ± 16.23 years. 59 (59%) were male and 41 (41%) were female. The median duration of hospital stay for the AIS patients was 7 days, with an interquartile range (IQR) of 4 to 16.5 days. The Mean National Institutes of Health Stroke Scale (NIHSS) at admission was 11.70 ± 7.57 points. The patients were classified according to the Trial of Org 10,172 in Acute Stroke Treatment (TOAST) classification system [12]. The distribution among the subcategories was as follows: Majority had Large-artery atherosclerosis (LAA): 44 patients (44%), followed by Cardio-embolism (CE): 29 patients (29%), Small-vessel occlusion (SVO): 13 patients, Stroke of other determined aetiology (SODE): 1 patient and Stroke of undetermined aetiology (SUC): 13 patients. Among the AIS cohort, 20 (20%) patients experienced a wake-up stroke. Several risk factors were found to be associated with AIS patients. A majority 55 (55%) had hypertension, followed by Diabetes Mellitus in 40 (40%) patients, dyslipidaemia in 4 (4%) patients, Coronary Artery

Disease (CAD) in 10 (10%) patients and 5 (5%) patients had a family history of stroke. Atrial Fibrillation was observed in 22 (22%) patients, with 12 (54%) among them having non-valvular atrial fibrillation. 10 (10%) patients had Rheumatic Heart Disease (Supplementary Table S1).

Current smoking/tobacco use, alcohol consumption, and abuse of recreational drugs was noted in 26 (26%), 17(17%) and 1 (1%) patient, respectively. A total of 20 patients (20%) received thrombolysis, at a mean duration of 193.75 ± 63.13 min from stroke onset.

Among them, 8 patients received alteplase, and 12 patients received tenecteplase. Four patients underwent mechanical thrombectomy. Six (6%) patients in the AIS cohort had in-hospital mortality. The median mRS at the time of admission was 4 with an IQR of (3,5), while at 3 months the median mRS was 3 with an IQR of (1,5)

The average age of the patients with haemorrhagic stroke was 55.72 ± 14.06 years. 68 (68%) were male and 32 (32%) were female. The median duration of hospital stay was 8 days.

with an interquartile range (IQR) of (5, 17.5) days. The majority 83 (83%) had hypertension, followed by Diabetes Mellitus in 19 (19%) patients, and Coronary Artery Disease (CAD) in 4 (4%) patients and 2 (2%) patients had a family history of stroke (Supplementary Table S2).

Current smoking/tobacco use, and alcohol consumption was noted in 26 (26%), and 21(21%) patients, respectively. 13 (13%) patients in the AIS cohort had in-hospital mortality. 13 (13%) patients in the ICH cohort had in-hospital mortality. The median mRS at the time of admission was 4 with an IQR of (2,5), while at 3 months the median mRS was 3 with an IQR of (2,5).

4 Results

A significant agreement in the total ASPECT score between the Raters 1 and 2 ($\kappa = 0.94(0.906, 0.973)$), Raters 1 and 3($\kappa = 0.93(0.896, 0.961)$) and Rater 2 and 3($\kappa = 0.93(0.887, 0.957)$) was noted. A good agreement was ascertained between Rater 1 and 3, even though the devices used were different (Fig. 2a and Table 1). A significant agreement in the binary ASPECT score (categorised as 0–5 and 6–10) between the Raters 1 and 2 ($\kappa = 0.95(0.870, 1.000)$), Rater 1 and 3 ($\kappa = 0.92(0.824, 1.000)$) and Raters 2 and 3 ($\kappa = 0.92 (0.829, 1.000)$) was noted. Regarding the various sub-components of ASPECTS, good to excellent inter-rater reliability was noted among all these Raters, across all the devices (Fig. 3a-e and Table 1). The highest inter-rater agreement was noticed in assessing the lentiform and insula regions ($\kappa = 0.81$ to 0.97) and the lowest agreement was noticed in assessing the M2, M3 and M6 regions ($\kappa = 0.68$ to 0.78).
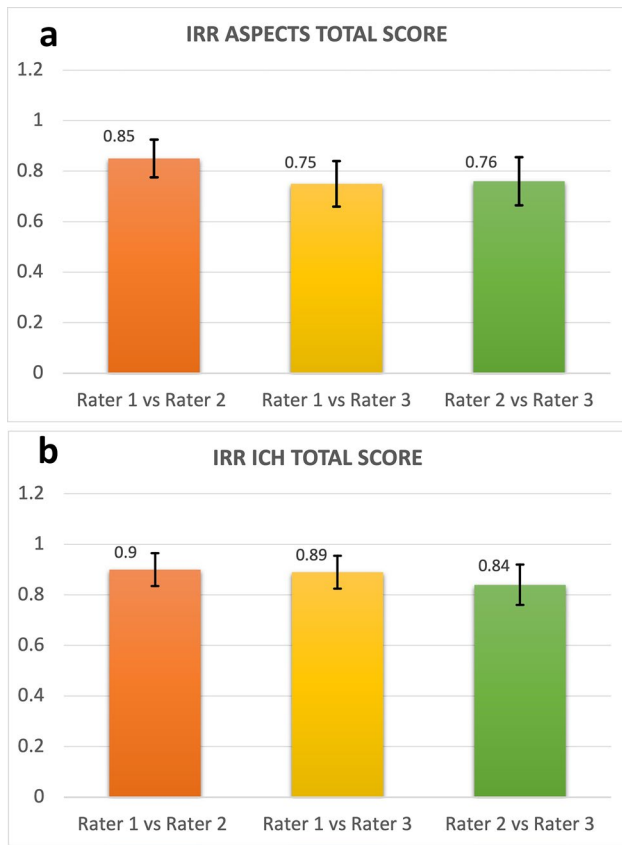
Fig. 2 Shows the inter-rater reliability (IRR) in the total scores of ASPECTS and ICH score

A kappa statistic was also obtained for a subset of patients based on severity of stroke. NIHSS (National Institute of Health Stroke Severity) score of 1–4, 5–15 and > 15 was considered minor, moderate and severe stroke with 14, 56 and 30 subjects in each subset. Among those with mild stroke (n = 14) Kappa coefficient for Rater 1 and 2 was noted to be nearly perfect (κ = 1.000) with undefined confidence interval which possibly happened because of the smaller number of individuals in the mild stroke category. The IRR between Rater 1 and 2 among those with moderate (n = 56) (κ = 0.89(0.812,0.962) and severe stroke (n = 30) and (κ = 0.963(0.903,1.000) were noted to be excellent. The IRR between Rater 1 and Rater 3 was noted to be (κ = 0.93(0.596,1.000), (κ = 0.92(0.868,0.963), and (κ = 0.91(0.858,0.968) for minor, moderate and severe stroke cases respectively. The IRR between Rater 2 and 3 noted to be (κ = 0.93(0.672,1.000), (κ = 0.92(0.851,0.965), and (κ = 0.91(0.857,0.981) for minor, moderate and severe stroke cases respectively.

The IRR among Raters 1 and 2 was noted to be good among the subjects with (n = 40) (κ = 0.93(0.869–0.978)) and without (n = 60) (κ = 0.95(0.897,0.990) diabetes. The IRR among Raters 1 and 3 was (κ = 0.94(0.884,0.981)) and (κ = 0.93(0.876, 0.960) with and without diabetes and that among the Raters 2 and 3 was (κ = 0.94(0.897,0.960) and (κ = 0.92(0.868, 0.960) with and without diabetes respectively. No linear relationship was noted between Age and

Table 1 Shows the inter-rater reliability between various Raters with regards to the ASPECTS score

Raters	Rater 1 vs Rater 2		Rater 1 vs Rater 3		Rater 2 vs Rater 3	
Sub-components of ASPECTS score	IRR	95% CI	IRR	95% CI	IRR	95% CI
Caudate	0.93	0.85–1.02	0.85	0.72–0.98	0.79	0.65–0.94
IC	0.92	0.83–1.01	0.76	0.69–0.91	0.69	0.51–0.87
L	0.86	0.75–0.97	0.90	0.81–0.99	0.81	0.69–0.94
Insula	0.97	0.92–1.02	0.92	0.84–1.011	0.90	0.81–0.99
M1	0.86	0.76–0.96	0.86	0.75–0.97	0.82	0.70–0.94
M2	0.85	0.74–0.96	0.78	0.65–0.92	0.68	0.52–0.84
M3	0.82	0.66–0.98	0.71	0.52–0.91	0.75	0.57–0.93
M4	0.90	0.81–0.99	0.80	0.67–0.93	0.81	0.68–0.94
M5	0.92	0.83–1.01	0.85	0.73–0.96	0.87	0.77–0.98
M6	0.91	0.811–1.01	0.77	0.62–0.92	0.69	0.52–0.86
TOTAL SCORE FOR ANTERIOR ASPECTS	0.85	0.77–0.92	0.75	0.67–0.85	0.76	0.67–0.86

M1, M2, M3—MCA cortex at level of basal ganglia, M4, M5 and M6—MCA cortex above the level of basal ganglia

ASPECTS Alberta Stroke Program Early CT Score, *IRR* Inter-rater reliability, *C* Caudate, *L* Lentiform, *IC* Internal Capsule
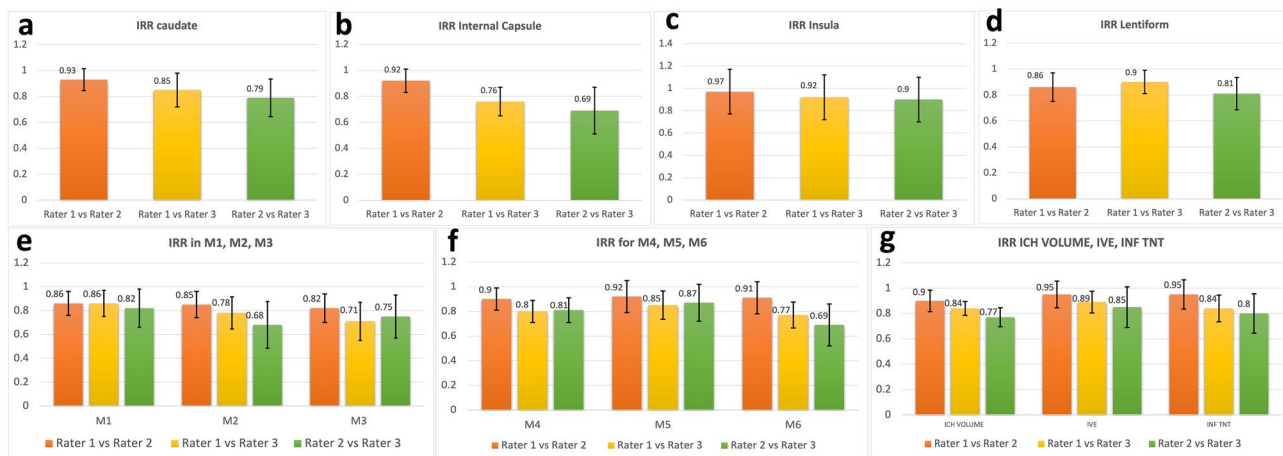
Fig. 3 Shows the inter-rater reliability (IRR) in the various sub-components of ASPECTS and ICH score

the ASPECT scores by each Raters 1, 2 and 3 (correlation coefficient -0.06, -0.03 and -0.05 respectively).

The kappa statistics for agreement in ICH score were noted to be excellent between Rater 1 and 2 (0.90(0.833, 0.965), Rater 1 and 3(0.90(0.832, 0.965) and Rater 2 and 3(0.85(0.769, 0.929) (Fig. 2b and Table 2). Similar to ASPECT score, the various sub-components of ICH score showed good to excellent inter-rater reliability was noted among all these Raters, across all the devices (Fig. 3f and Table 2). The highest inter-rater agreement was noticed in assessing the IVE (Intraventricular Extension) and infratentorial extension IFN TNT (κ = 0.95 to 0.80) and the lowest agreement was noticed in assessing the ICH volume (κ = 0.77).

5 Discussion

Inter-rater reliability in ASPECTS (Alberta Stroke Program Early CT Score) is important while guiding stroke treatment because it helps ensure that the score is calculated consistently and accurately by different Raters. In APSECTS

scoring, the middle cerebral artery territory is partitioned into ten distinct regions, with a one-point deduction made for areas displaying early ischemic indicators, such as focal swelling or parenchymal hypoattenuation. The subcortical structures are allotted 3 points (Caudate, Lentiform, and Internal Capsule), while the MCA cortex is allotted 7 points (insular cortex, M1, M2, M3 at level of basal ganglia, M4, M5 and M6 above it) [4]. Consistency and agreement among healthcare professionals in ASPECTS scoring are crucial to ensure accurate evaluation of the imaging findings to guide treatment decisions effectively and avoid discrepancies based on the severity of ischemic lesions [13].

Since its introduction in 2000 by Barber et al [5] various studies have examined the inter-rater agreement of ASPECTS scoring among various healthcare professionals. In the present study, significant agreement in the total ASPECT score between the Raters 1 and 2 (κ = 0.94(0.906,0.973)), Raters 1 and 3(κ = 0.93(0.896,0.961)) and Rater 2 and 3(κ = 0.93(0.887,0.957)) was noted, in line with previous studies. The Calgary group conducted an early study involving six physicians, including three non-neuroradiologists and three neuroradiologists, which reported a high level of

Table 2 Shows the inter-rater reliability between various Raters with regard to the ICH score

Raters	Rater 1 vs Rater 2		Rater 1 vs Rater 3		Rater 2 vs Rater 3	
Sub-components of ASPECTS score	IRR	95% CI	IRR	95% CI	IRR	95% CI
ICH Vol	0.9	0.81–0.98	0.84	0.73–0.94	0.77	0.66–0.89
IVE	0.95	0.90–1.01	0.89	0.81–0.98	0.85	0.75–0.96
INF TNT	0.95	0.89–1.04	0.84	0.68–1.00	0.80	0.65—0.96
ICH score	0.90	0.83–0.96	0.89	0.83–0.96	0.84	0.76–0.92

ICH Intra-cerebral haemorrhage, *IRR* Inter-rater reliability, *CI* confidence interval, *ICH vol* ICH volume, *IVE* intra-ventricular extension, *INF TNT* Infratentorial origin of hemorrhage

disagreement [14]. Another study by Gupta et al. demonstrated moderate interrater agreement (κ = 0.53, Cohen's κ) between two experienced neuroradiologists for 155 patients [15]. In contrast, Finlayson et al. showed very good agreement among four readers, consisting of two neuroradiologists and two neurologists, for 181 patients, using Cronbach's α to assess internal consistency (α = 0.83) [16]. Additionally, Nicholson et al. recently reported a relatively good overall agreement (κ = 0.66, Cohen's κ) at an ASPECTS threshold of 5 [17]. Nimmalapudi et al., assessed the inter-rater reliability between emergency medicine physicians and radiologists in grading ASPECTS scores among patients with acute ischemic stroke. Even in a rural setting, there was significant agreement between the emergency medicine physician and the radiologist when grading ASPECTS [18].

Regarding the per-region assessment of ASPECTS, in this study, the greatest inter-rater agreement was noticed in assessing the lentiform and insula regions and the lowest agreement was noticed in assessing the internal capsule, M2, M3 and M6 regions. The best agreement was noticed for Insula (0.90 to 0.97). Nicholson et al. quantified per-region agreement for ASPECTS on NCCT brain images from 375 patients with acute ischemic stroke, as assessed by two experienced neuroradiologists. In their dataset, the greatest agreement was found in the caudate, lentiform, and M5 regions, with the M3 and internal capsule regions showing the least interobserver agreement [17].

In the study by Gupta et al., per-region assessment also showed variability between two neuroradiologists varying from κ = 0.73 in the insula and 0.71 in the caudate to 0.28 in the internal capsule, although this was attributed to the very small number of internal capsule strokes in their sample [15].

Finlayson et al. found that the insula, lentiform and caudate exhibited the highest correlation coefficient at Cronbach's α = 0.71 to 0.79, while the internal capsule demonstrated the lowest correlation at α = 0.37 [16]. Chu et al., evaluating the interrater agreement for ASPECTS on 91 acute ischemic patients, also found that the lowest concordance was seen in the internal capsule on NCCT [19].

In most of these including our study, the area with the lowest agreement is the internal capsule. There can be various physiological and methodological reasons for this. Various regions of the brain exhibit different degrees of sensitivity to ischemic changes. In cases where the caudate and lentiform are affected by infarction, it is common for the anterior limb of the internal capsule to be involved due to a shared blood supply. Nevertheless, some assessors have specifically identified ischemic changes using only the posterior limb of the internal capsule [14].

The internal capsule frequently appears relatively hypoattenuated on NCCT brain scans in healthy individuals, which can pose challenges in identifying additional hypoattenuation associated with early ischemia. Moreover, the initial paper from the Calgary group in 2001, introducing the ASPECTS score, documented disagreements among the six physicians who developed ASPECTS regarding the definition of the internal capsule for scoring purposes [14]. Readers varied in their assessment approach, with some evaluating both limbs of the internal capsule and deducting a point if either limb was affected, while others focused solely on the posterior limb. Given the initial disagreement among the group that developed the ASPECTS score, it is understandable that discrepancies also exist among less experienced readers. This is particularly true considering the highly variable blood supply to the internal capsule and its anterior and posterior limbs.

Additionally, in assessing the M1- M6 components, the lower to lowest agreements in various studies were seen in the M3 region [(M3 region, α = 0.39 (Horn et al.,), α = 0.69 (Finlayson et al.,) κ = 0.34 (Nicholson et al.,)]. In our study also the M3 and M2 regions had lower agreement compared to other M1-M6 components.

Within the territories of subganglionic and supraganglionic nuclei (M1–M6), identifying cortical features plays a significant role for early detection of infarcts. The proximity of these regions to the skull poses challenges, with beam hardening artifacts potentially complicating the differentiation process. Additionally, the M3 region exhibits a lower level of agreement, possibly attributed to its relatively infrequent occurrence [20]. The infrequency of M3 occurrences might contribute to the limited familiarity and consensus among Raters, leading to a decrease in inter-rater agreement. These factors highlight the complexity of assessing certain brain regions, emphasizing the need for subtle evaluation strategies and continued research to enhance reliability in the detection of early infarcts. As interventional treatment decisions are often based on the ASPECT Score, decision-makers should be aware of these limitations of this system.

On the contrary, Horn et al., Nicholson et al., and Gupta et al., reported the highest agreement for the insula region (κ = 0.96, α = 0.71, and κ = 0.73, respectively) [21, 17, 15], in line with our findings. This might be due to its distinct anatomical location, surrounded by the cerebrospinal fluid (CSF)-filled Sylvian fissure laterally, the operculum brain parenchyma laterally, and the relatively hypoattenuated external capsule medially [22]. Additionally, the susceptibility to ischemic changes caused by hypoperfusion varies across different brain regions, with the insular cortex, precentral gyrus, and basal ganglia being the most vulnerable [20].

Stroke severity is a strong determinant of outcomes in patients receiving mechanical thrombectomy (MT) in clinical trials, and heavily influences the decision of whether to perform an MT. The efficacy of MT in patients with severe strokes is clearly larger than among those with less severe deficits [23]. Though our findings are consistent with

previous studies that have reported high inter-rater reliability of ASPECTS [21, 16], it's important to note that these studies did not stratify their results based on stroke severity, which is a unique aspect of our study. In our study, the IRR was nearly perfect ($\kappa = 1.000$) for those with mild stroke, and excellent for those with moderate ($\kappa = 0.89$) and severe stroke ($\kappa = 0.963$) between Rater 1 and 2. Similarly, the IRR between Rater 1 and Rater 3, and Rater 2 and 3 were also excellent across all stroke severities. These findings provide valuable insights into the consistency of ASPECTS assessments across varying stroke presentations. However, it's worth noting that the confidence interval for the kappa coefficient in the mild stroke category was undefined, possibly due to the smaller sample size. Future research with a larger sample size may be needed to confirm these findings in patients with mild strokes.

The relationship between age and ASPECT scores has not been explicitly studied in literature. However, age is often considered a factor in stroke severity and outcomes and in clinical trials, age has been associated with worse outcomes in patients with large vessel occlusion (LVO), with or without MT [24, 23]. Given this, the lack of a significant correlation between age and ASPECT scores in our study could be seen as a neutral finding. It neither contradicts nor directly supports the existing literature. It does, however, add a new dimension to the understanding of these variables in stroke assessment and could be a valuable contribution to the field. This may be explored in future studies.

Diabetic patients are at increased risk for stroke, but little is known about the presence of other brain lesions. In a previous study Schmidt et al., demonstrated that patients with diabetes had more extensive cortical atrophy and a higher ventricle-to-brain ratio than in nondiabetic patients [25]. Van Harten et al., systematically reviewed brain imaging studies in patients with diabetes and showed a relation between diabetes and cerebral atrophy and lacunar infarcts but no consistent relation with White matter lesions [26]. In the Secondary Prevention of Small Subcortical Strokes (SPS3) trial, which included participants with recent lacunar infarcts, those with a history of diabetes had distinct neuroimaging characteristics on magnetic resonance imaging (MRI) as compared with those without diabetes, with an increased odds of posterior circulation infarcts and a lower burden of microbleeds and enlarged perivascular spaces. These factors can potentially affect ASPECTS assessment. Therefore in our study, we analyzed the IRR of ASPECTS between the Raters separately in patients with and without diabetes, which is another unique aspect of our study. The findings revealed moderate to good reliability across all Raters irrespective of the diabetes status.

Presently, both American and European guidelines strongly recommend endovascular treatment (EVT) for patients who do not show significant early ischemic changes on NCCT. based on an ASPECTS of 6 or higher. However, recent high-quality evidence from four randomized controlled clinical trials (RCTs) has emerged, demonstrating that compared to best medical therapy (BMT), EVT is significantly associated with reduced disability (OR 1.70, 95% CI 1.39 to 2.07), increased independent ambulation (risk ratio (RR) 1.69, 95% CI 1.33 to 2.14), and improved functional outcomes at 3 months post-treatment (RR 2.33, 95% CI 1.76 to 3.10). Although the rates of symptomatic intracranial hemorrhage (sICH) and any intracranial hemorrhage (ICH) were higher in the EVT group, there was no difference in 3-month mortality between the two groups (RR 0.98, 95% CI 0.83 to 1.15). Despite these promising results, the efficacy and safety of EVT in patients with low ASPECTS remain a grey area. Future trials are expected to continue exploring this aspect. In light of this, we dichotomized ASPECTS into two subgroups: low (5 or less) and high (greater than 5). Our study found a significant agreement in the binary ASPECT score (categorized as 0–5 and 6–10) between the raters. The kappa coefficients between Raters 1 and 2, Rater 1 and 3, and Raters 2 and 3 were 0.95, 0.92, and 0.92 respectively, indicating a high level of inter-rater reliability. In the context of the existing literature, our study's findings are consistent with those of Nicholson et al., who found relatively good agreement when ASPECTS were dichotomized into 0–5 versus 6–10 ($\kappa = 0.66$; 0.49 to 0.84). However, our study demonstrated a higher level of agreement, suggesting that the reliability of ASPECTS may be even greater than previously reported. On the other hand, Gupta et al. reported a moderate agreement for dichotomized baseline ASPECTS (< 7 versus > 7), with a kappa of 0.53. This discrepancy could be due to differences in the categorization of ASPECTS or variations in the patient populations and Raters' expertise.

All the studies discussed till now have focused on evaluating the inter-rater reliability among healthcare professionals who recorded their observations using traditional PACS or similar platforms. With the increasing adoption of smartphones and healthcare apps, there has been a shift towards utilizing these technologies to enhance patient care. Zeiger et al., assessed the current patterns of smartphone use and perceptions of its utility for patient care among academic neurology trainees and attending physicians. A questionnaire was developed and 213 responses were received. The findings revealed that neurology trainees used smartphones more frequently for patient care activities compared to attending physicians ($P = 0.005$) [28]. Various smart phone Apps have been developed and evaluated for their role in enhancing patent care in stroke patients.

Tsai et al assessed improvement in interhospital communication and reduction in door-to-puncture time in interhospital transfer patients undergoing thrombectomy using a smartphone application called "LINE." Retrospective data

analysis was conducted on transfer patients over three periods: before using the smartphone application (2017), the first year of using the application (2018), and the second year of using the application (2019). The results showed a significant decrease in the mean door-to-puncture time from 109 to 92 min over the study period. The adoption of the smartphone application facilitated hub-to-spoke communication and improved interhospital connections, leading to time savings [29].

Takao evaluated the impact of a smartphone application called JOIN on stroke care. JOIN facilitated real-time information exchange, tracking of patient management, and communication among the healthcare team. The study compared the period before implementing JOIN with the period after and reviewed data from 91 stroke patients treated with tPA and/or thrombectomy. DICOM images (e.g., computed tomography [CT], magnetic resonance imaging [MRI]) were acquired and shared via the smartphone interface, and the decisions with the related images were evaluated. The results showed significant reductions in patient management time and high satisfaction among staff members. Patient outcomes and treatment costs remained unchanged [30]. On the other hand, Monsour et al. evaluated the performance of automated ASPECTS compared to ASPECTS interpretations of CT images sent on WhatsApp. The study included 122 patients with anterior circulation stroke who received intravenous thrombolysis. All transferred NCCT images were captured by one universal smartphone—the "resident's phone" (Iphone 8 plus; 8MP camera). It was found that the use of WhatsApp captured CT images for ASPECTS assessment resulted in a significant reduction in the performance of automated ASPECTS provided by the RAPID@ IschemaView ASPECTS system [31]. Therefore, given the growing interest in using smartphone apps in healthcare and the concern of reduced performance for ASPECTS when devices or platforms are different, it is crucial to further assess the level of agreement among Raters using different devices and platforms.

India is one of the largest and fastest-growing markets for digital consumers with more than half a billion internet subscribers [32]. India is digitizing faster than many mature and emerging economies. India had 560 million internet subscribers in September 2018, second only to China [32]. The cost of mobile data in India is one of the cheapest in the world. In fact, at Rs 6.7 ($0.09) per gigabyte (GB), the average cost of mobile data in India is the cheapest in the world [33, 34, 35]. In 2023, India currently holds the position of the third lowest data rate at USD 0.17 (~ Rs 13.98) [36]. The SMART INDIA App aims to help district-level clinicians in such decision-making by testing the efficacy of two innovative solutions including, firstly by using a low-cost Tele-stroke model and second by training physicians in district hospitals to diagnose and manage acute stroke

('Stroke physician model') [11]. The SMART INDIA App underpins the concept of seamless communication and coordination between healthcare professionals involved in acute stroke care. The workflow in the SMART INDIA App mandates that once an acute stroke patient arrives at the designated district hospital, the attending physician at the hospital immediately logs into the App and starts entering the details of the stroke patient, capturing the relevant information for further assessment and management. The Brain CT images are also uploaded as a video-recorded file to the App. Once the patient's information is entered, neurologists receive a notification alert on their smartphones through the SMART INDIA App. This notification serves as an immediate alert to the neurologist that a stroke case requires their expertise and input. The neurologist and the attending physician engage in real-time communication using the chat box feature within the SMART INDIA App. They discuss the patient's condition, review diagnostic information including the CT scans of the patient uploaded in the app, and collaborate on specific management decisions. Based on the discussion and assessment, the neurologist and physician jointly determine the appropriate management plan for the acute stroke patient. They may decide on interventions, medication, or further diagnostic tests as needed [11]. Hence, it is important to ensure that the neurologists reading the recorded images in the App demonstrate a level of reliability comparable to that of neuroradiologists reading the same images through a workstation console utilizing Picture Archiving and Communication Systems (PACS).

By validating the reliability of neurologists' interpretations, we can ascertain whether the recommended treatment provided to patients is appropriate and in line with established standards. The results of the present study showed a significant agreement among all the Raters in terms of total ASPECT score (0.76 to 0.85) and also with regards to the various subcomponents of ASPECTS (ranging from 0.68 to 0.97). The findings of this study are consistent with previous studies conducted on a similar hypothesis. Sakai et al. found favorable inter-device agreement in the assessment of DWI-ASPECTS (Diffusion-Weighted Imaging-Alberta Stroke Program Early CT Score) between a smartphone and a desktop PC monitor. Two vascular neurologists participated in the study, and both demonstrated high levels of agreement with the neuroimaging findings when comparing the smartphone and PC monitor ($\kappa = 0.777$ and 0.787). The interrater agreement between the neurologists was also reported to be high ($\kappa = 0.710$ and $\kappa = 0.663$) [37]. In another study, Brehm et al., examined the use of mRay software on handheld devices (iPhone 7 Plus and MED-TAB) for evaluating CT scans of patients with suspected stroke. They found that the mobile devices detected all major abnormalities, including large-vessel occlusions and intracranial haemorrhages. There was no significant difference in the detection of ischemic signs (ASPECTS) and other

stroke-related features between the handheld devices and a traditional workstation. Both Raters considered the diagnostic quality of the handheld devices to be sufficient for making treatment decisions [38].

Similarly for the ICH score, the kappa statistics for agreement in ICH score were noted to be excellent between all the Raters, even using different devices and platforms (0.85 to 0.90). The inter-rater reliability could not be calculated for the sub-components of GCS and age, as these variables were fixed values provided to the raters in advance. Among the rest three sub-components (ICH volume, intra-ventricular extension, and Infratentorial origin of haemorrhage) the inter-rater reliability was also good to excellent ranging from 0.77 to 0.95. These findings are consistent with previous studies conducted on a similar hypothesis.

Dowlatshahi et al., evaluated the reliability of five NCCT markers used to predict hematoma expansion in intracerebral haemorrhage patients. Four independent readers analyzed images from 40 patients and assessed the presence or absence of intra-hematoma hypodensities, blend sign, fluid level, irregular hematoma morphology, and heterogeneous hematoma density. The study found that there was excellent interrater agreement among the readers for all five markers, with agreement ranging from 94 to 98%. The interrater kappa values ranged from 0.67 to 0.91, with the lowest value observed for fluid level. Similarly, interrater agreement showed a consistent pattern, ranging from 89 to 93%, with kappa values ranging from 0.60 to 0.89 [39]. However, no specific studies have yet evaluated the inter-rater reliability of the ICH score using different devices and/ or platforms, which was assessed in this study and was also found to have excellent reliability.

6 Limitations

A total of three Raters participated in the assessment of the scores. Future studies may involve a larger number of Raters to enhance the reliability and accuracy of the assessments. Also, the study was conducted in a controlled setting in a single centre and may not be generalizable to all clinical settings. Lastly, some of the readings performed by the neurologists using their personal mobile phones could have been assessed under suboptimal environmental and lighting conditions.

7 Conclusion

The results of the study suggest that the neurologists estimating the ASPECTS and ICH score had good to excellent agreement with the neuro-radiologist (gold standard), even though the devices and platforms used were different. Therefore, from neuroimaging perspective, the interpretations of a neurologist from the SMART INDIA App can be considered

as valid and reliable. These results contribute to our understanding of the feasibility and reliability of app-based image evaluation in the assessment of strokes.

Declarations

References

1. Kenda M, Cheng Z, Guettler C, et al. Inter-rater agreement between humans and computer in quantitative assessment of computed tomography after cardiac arrest. Front Neurol. 2022;13. Accessed May 20, 2023, from https://doi.org/10.3389/fneur.2022.990208.
2. Cole R. Inter-Rater Reliability Methods in Qualitative Case Study Research. Sociol Methods Res. 2023;22:00491241231156971.
3. Galinovic I, Puig J, Neeb L, Guibernau J, Kemmling A, Siemonsen S, et al. Visual and Region of Interest-Based

Inter-Rater Agreement in the Assessment of the Diffusion-Weighted Imaging– Fluid-Attenuated Inversion Recovery Mismatch. Stroke. 2014;45(4):1170–2.

4. Mokin M, Primiani CT, Siddiqui AH, Turk AS. ASPECTS (Alberta Stroke Program Early CT Score) Measurement Using Hounsfield Unit Values When Selecting Patients for Stroke Thrombectomy. Stroke. 2017;48(6):1574–9.

5. Barber PA, Demchuk AM, Zhang J, Buchan AM. Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolyt. ASPECTS Study Group. Alberta Stroke Programme Early CT Score. Lancet Lond Engl. 2000;13;355(9216):1670–4.

6. Yoo AJ, Zaidat OO, Chaudhry ZA, Berkhemer OA, González RG, Goyal M, et al. Impact of Pretreatment Noncontrast CT Alberta Stroke Program Early CT Score on Clinical Outcome After Intra-Arterial Stroke Therapy. Stroke. 2014;45(3):746–51.

7. Yoo AJ, Berkhemer OA, Fransen PSS, van den Berg LA, Beumer D, Lingsma HF, et al. Effect of baseline Alberta Stroke Program Early CT Score on safety and efficacy of intra-arterial treatment: a subgroup analysis of a randomised phase 3 trial (MR CLEAN). Lancet Neurol. 2016;15(7):685–94.

8. Powers WJ, Derdeyn CP, Biller J, Coffey CS, Hoh BL, Jauch EC, et al. 2015 American Heart Association/American Stroke Association Focused Update of the 2013 Guidelines for the Early Management of Patients With Acute Ischemic Stroke Regarding Endovascular Treatment. Stroke. 2015;46(10):3020–35.

9. Hemphill JC, Bonovich DC, Besmertis L, Manley GT, Johnston SC. The ICH Score. Stroke. 2001;32(4):891–7.

10. Hemphill JC, Greenberg SM, Anderson CS, Becker K, Bendok BR, Cushman M, et al. Guidelines for the Management of Spontaneous Intracerebral Hemorrhage. Stroke. 2015;46(7):2032–60.

11. Vishnu VY, Bhatia R, Khurana D, Ray S, Sharma S, Kulkarni GB, et al. Smartphone-Based Telestroke Vs"Stroke Physician" led Acute Stroke Management (SMART INDIA): A Protocol for a Cluster-Randomized Trial. Ann Indian Acad Neurol. 2022;25(3):422–7.

12. Adams HP, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, et al. Classification of subtype of acute isc. Stroke. 1993;24(1):2–3.

13. Bing F, Berger I, Fabry A, Moroni AL, Casile C, Morel N, et al. Intra- and inter-rater consistency of dual assessment by radiologist and neurologist for evaluating DWI-ASPECTS in ischemic stroke. Rev Neurol (Paris). 2022;178(3):219–25.

14. Pexman JH, Barber PA, Hill MD, Sevick RJ, Demchuk AM, Hudon ME, et al. Use of the Alberta Stroke Program Early CT Score (ASPECTS) for assessing CT scans in patients with acute stroke. AJNR Am J Neuroradiol. 2001;22(8):1534–42.

15. Gupta AC, Schaefer PW, Chaudhry ZA, Leslie-Mazwi TM, Chandra RV, González RG, et al. Interobserver reliability of baseline noncontrast CT Alberta Stroke Program Early CT Score for intra-arterial stroke treatment selection. AJNR Am J Neuroradiol. 2012;33(6):1046–9.

16. Finlayson O, John V, Yeung R, Dowlatshahi D, Howard P, Zhang L, et al. Interobserver Agreement of ASPECT Score Distribution for Noncontrast CT, CT Angiography, and CT Perfusion in Acute Stroke. Stroke. 2013;44(1):234–6.

17. Nicholson P, Hilditch CA, Neuhaus A, Seyedsaadat SM, Benson JC, Mark I, et al. Per-region interobserver agreement of Alberta Stroke Program Early CT Scores (ASPECTS). J NeuroInterventional Surg. 2020;12(11):1069–71.

18. Nimmalapudi S, Inampudi V, Prakash A, Gowda R, Varadharajan S. Understanding ASPECTS of stroke: Inter-rater reliability between emergency medicine physician and radiologist in a rural setup. Neuroradiol J. 2022;31:197140092211144.

19. Chu Y, Ma G, Xu XQ, Lu SS, Cao YZ, Shi HB, et al. Total and regional ASPECT score for non-contrast CT, CT angiography, and CT perfusion: inter-rater agreement and its association with the final infarction in acute ischemic stroke. Acta Radiol. 2022;63(8):1093–101.

20. Payabvash S, Souza LCS, Wang Y, Schaefer PW, Furie KL, Halpern EF, et al. Regional Ischemic Vulnerability of the Brain to Hypoperfusion. Stroke. 2011;42(5):1255–60.

21. Van Horn N, Kniep H, Broocks G, Meyer L, Flottmann F, Bechstein M, et al. ASPECTS Interobserver Agreement of 100 Investigators from the TENSION Study. Clin Neuroradiol. 2021;31(4):1093–100.

22. Kortz MW, Lillehei KO. Insular Cortex. In: StatPearls. StatPearls Publishing. 2023. Accessed November 19, 2023, from http://www.ncbi.nlm.nih.gov/books/NBK570606/.

23. Bres-Bullrich M, Fridman S, Sposato LA. Relative Effect of Stroke Severity and Age on Outcomes of Mechanical Thrombectomy in Acute Ischemic Stroke. Stroke. 2021;52(9):2846–8. https://doi.org/10.1161/STROKEAHA.121.034946.

24. Hill MD, Demchuk AM, Tomsick TA, Palesch YY, Broderick JP. Using the Baseline CT Scan to Select AcuteStroke Patients for IV-IA Therapy. Am J Neuroradiol. 2006;27(8):1612–6.

25. Schmidt R, Launer LJ, Nilsson LG, Pajak A, Sans S, Berger K, et al. Magnetic Resonance Imaging of the Brain in Diabetes. Diabetes. 2004;53(3):687–92.

26. Van Harten B, De Leeuw FE, Weinstein HC, Scheltens P, Biessels GJ. Brain Imaging in Patients With Diabetes. Diabetes Care. 2006;29(11):2539–48.

27. Kappelhof M, Jansen IGH, Ospel JM, Yoo AJ, Beenen LFM, Roosendaal SD, et al. Endovascular Treatment May Benefit Patients With Low Baseline Alberta Stroke Program Early CT Score: Results From the MR CLEAN Registry. Stroke Vasc Interv Neurol. 2022;2(3): e000199.

28. Zeiger W, DeBoer S, Probasco J. Patterns and Perceptions of Smartphone Use Among Academic Neurologists in the United States: Questionnaire Survey. JMIR MHealth UHealth. 2020;8(12): e22792.

29. Tsai ST, Wang WC, Lin YT, et al. Use of a Smartphone Application to Speed Up Interhospital Transfer of Acute Ischemic Stroke Patients for Thrombectomy. Front Neurol. 2021;12. Accessed May 23, 2023, from https://www.frontiersin.org/articles/10.3389/fneur.2021.606673.

30. Takao H, Sakai K, Mitsumura H, et al. A Smartphone Application as a Telemedicine Tool for Stroke CareManagement. Neurol Med Chir (Tokyo). 2021;61(4):260–7. https://doi.org/10.2176/nmc.oa.2020-0302.

31. Mansour OY, Ramadan I, Abdo A, Hamdi M, Eldeeb H, Marouf H, et al. Deciding Thrombolysis in AIS Based on Automated versus on WhatsApp Interpreted ASPECTS, a Reliability and Cost-Effectiveness Analysis in Developing System of Care. Front Neurol. 2020;11:333.

32. Digital India: Technology to transform a connected nation | McKinsey. Accessed July 4, 2023, from https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/digital-india-technology-to-transform-a-connected-nation.

33. Khan S. India ranks lowest on global mobile data cost and that's great news [here's how]. Published July 26, 2021. Accessed July 4, 2023, from https://www.ibtimes.co.in/india-ranks-lowest-global-mobile-data-cost-thats-great-news-heres-how-839072.

34. Mobile data: Why India has the world's cheapest. BBC News. https://www.bbc.com/news/world-asia-india-47537201. Published March 18, 2019. Accessed July 4, 2023.

35. Mobile data in India is the third cheapest in the world. You will be SHOCKED to know where it is costliest. Accessed January 21, 2024. https://www.firstpost.com/world/india-has-the-third-cheapest-mobile-data-in-the-world-people-is-us-pay-33x-more-than-indians-12614602.html.

36. At $0.09/GB, India's data plans cheapest. The Times of India. https://timesofindia.indiatimes.com/business/india-business/at-0-09/gb-indias-data-planscheapest/articleshow/77728709.cms.

37. Sakai K, Komatsu T, Iguchi Y, Takao H, Ishibashi T, Murayama Y. Reliability of Smartphone for Diffusion-Weighted Imaging-Alberta Stroke Program Early Computed Tomography Scores in Acute Ischemic Stroke Patients: Diagnostic Test Accuracy Study. J Med Internet Res. 2020;22(6): e15893.
38. Brehm A, Maus V, Khadhraoui E, Psychogios MN. Image review on mobile devices for suspected stroke patients: Evaluation of the mRay software solution. PLoS ONE. 2019;14(6): e0219051.
39. Dowlatshahi D, Morotti A, Al-Ajlan FS, Boulouis G, Warren AD, Petrcich W, et al. Interrater and Intrarater Measurement Reliability of Noncontrast Computed Tomography Predictors of Intracerebral Hemorrhage Expansion. Stroke. 2019;50(5):1260–2.