ENHANCING EFFICIENCY IN LOW-RISK CHEST X-RAY REPORTING: A COMPARATIVE STUDY OF MANUAL, TEMPLATE-BASED, AND AI-GENERATED METHODS

Marek Řehoř, Šimon Kličník, Jakub Dandár, Daniel Kvak

Abstract

Efficient and accurate chest X-ray (CXR) reporting is essential in radiology, especially for quickly identifying low-risk cases to prioritize more complex ones. This study investigates the time efficiency of three CXR reporting methods: manual, template-based, and AI-generated, focusing specifically on low-risk CXR evaluations in a radiology department. Results show that manual reporting, which requires free-text documentation, takes significantly longer than other methods, with average mean times per study of 96.4 seconds (RAD1), 91 seconds (RAD2), and 70.8 seconds (RAD3). In contrast, the structured, template-based approach reduced these times to 32.9 seconds (RAD1), 32 seconds (RAD2), and 48.8 seconds (RAD3), representing an average efficiency improvement of 53.93% compared to manual reporting. The AI-generated method yielded the shortest mean times per study at 27.7 seconds (RAD1), 31.9 seconds (RAD2), and 33.8 seconds (RAD3), with an average reduction of 62.82% compared to manual reporting. In conclusion, AI-generated reporting offers substantial time savings and maintains high accuracy, indicating strong potential to enhance radiology workflow efficiency. This study supports the integration of AI into routine CXR reporting, enabling radiologists to focus more on complex cases. Future research should explore the long-term impacts and further improvement of AI algorithms to optimize radiology practices.

Keywords

Artificial Intelligence, Chest X-ray Reporting, Diagnostic Reporting Methods, Radiology, Time Efficiency, Workflow Efficiency

1 Introduction

Chest X-ray (CXR) imaging remains one of the most commonly performed radiological procedures worldwide, particularly for evaluating thoracic and cardiac diseases [1]. As healthcare systems face increased patient loads, radiology departments are challenged with maintaining timely and accurate diagnostic reporting. Low-risk CXRs are defined as studies that show no significant abnormalities or findings of clinical concern, such as normal pulmonary, cardiac, and mediastinal appearances, and are often encountered during routine screenings or follow-up evaluations. Efficient reporting of low-risk CXRs is particularly essential, as it allows radiologists to prioritize more complex cases with significant abnormalities, ultimately optimizing workflow and improving patient outcomes [2].

In recent years, artificial intelligence (AI) has shown promise in augmenting radiologists' efficiency and accuracy through computer-aided detection (CAD) systems [3]. Specifically, deep learning algorithms have advanced significantly, offering automated tools that can detect and localize abnormalities in CXR images with comparable accuracy to human experts [4–7]. These systems utilize convolutional neural networks (CNNs) to identify a variety of thoracic pathologies. By providing preliminary interpretations, AI-based systems can assist radiologists in expediting low-risk evaluations and focusing on high-complexity cases.

This study aims to evaluate the time efficiency of AI-generated reporting in comparison with traditional manual and template-based methods for low-risk CXRs. Specifically, we examine whether the use of AI-generated preliminary reports offers measurable time savings that could enable radiologists to allocate more attention to complex cases. Through a comparative analysis involving experienced radiologists, this study seeks to provide insights into the operational impact of AI on routine radiology workflows, contributing to the broader understanding of AI's role in enhancing healthcare efficiency.

2 Materials and Methods

2.1 Study Design

The study was designed as a controlled, single-center, comparative analysis aimed at evaluating the time efficiency of three different reporting methods—manual, template-based, and AI-generated reporting—for interpreting low-risk chest X-rays (CXR). The primary objective was to determine whether AI-generated reports could provide significant time savings without compromising accuracy, thereby enhancing workflow efficiency in routine, low-risk CXR evaluations. A crossover design was selected to allow each participating radiologist to engage with all three reporting methods, enabling direct comparison while controlling for potential learning effects. The controlled environment of this design was chosen to produce reliable insights into how each method impacts reporting time and accuracy under consistent conditions, using a representative set of low-risk CXR cases.

2.2 Sample

Data for this study were obtained from a specialized center, OTRAN Kutná Hora, which provides comprehensive care for patients with chronic lung diseases, including preventive

Figure 1 – Examples of low-risk chest X-rays used in the study.

screenings, diagnostic assessments, and treatment options. The center serves a diverse patient population with common conditions such as chronic obstructive pulmonary disease (COPD) and bronchial asthma, as well as rarer lung diseases. Routine evaluations are also performed for patients presenting with symptoms like persistent cough, mucus production, chest wheezing, dyspnea, and chest pain.

For the purposes of this study, a randomly selected sample of 30 anonymized CXRs, performed between October 1st and October 14th, 2024, was acquired to represent low-risk cases commonly encountered in clinical practice (Figure 1). Demographic information for this sample includes patient age, sex, and the imaging equipment used. The mean age of patients in this dataset is 57.5±16.3 years, with an age range from 19 to 83 years (Table 1). The sample consists of 19 females (63%) and 11 males (37%). X-ray imaging was conducted using equipment from various manufacturers and models. The DRGEM Diamond model was the most frequently used (n=20), followed by Canon Inc. CXDI Control Software NE (n=7). Additionally, individual cases were imaged using Siemens Fluorospot Compact FD (n=1), GE Healthcare Discovery XR656 (n=1), and Samsung Electronics GR40CWC (n=1).

Attribute	Female (n=19)	Male (n=11)
Mean Age ± SD	62.6 ± 12.8	48.7 ± 18.5
Age Range	41–83	19–80

Table 1 – Age and sex distribution of patients.

2.3 Assessment

This study assessed the reporting efficiency and accuracy of three methods for interpreting low-risk CXR: manual reporting, template-based reporting, and AI-generated reporting.

- **Manual Reporting**: Radiologists performed standard manual reporting, where they reviewed each CXR and documented their findings in free-text format. This approach followed departmental guidelines for documenting observations in a non-structured, descriptive manner, as is customary in routine clinical practice.
- **Template-Based Reporting**: Radiologists used a structured reporting template designed to streamline the documentation of low-risk CXR evaluations. The template included predefined fields covering essential descriptors, promoting consistency and speed in report creation while reducing variability in wording.
- **AI-Generated Reporting**: A commercial-stage, MDR-certified computer-aided detection software, Carebot AI CXR v2 (Figure 2), was used to generate preliminary interpretations for each assigned CXR to identify and localize pulmonary abnormalities.

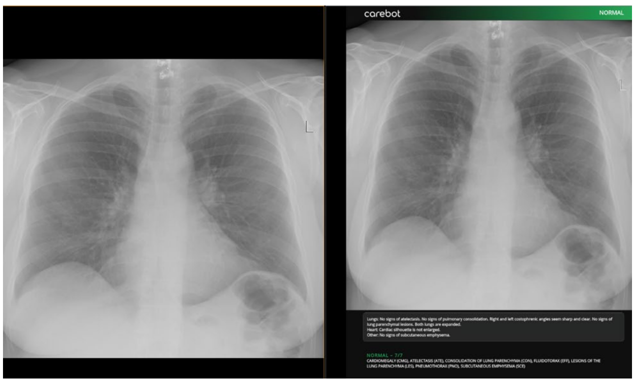


Figure 2 – An example of the examined AI model (Carebot AI CXR v2; Carebot s.r.o.) predictions, presented in DICOM viewer (Weasis Medical Viewer v4.4.0).

Three independent radiologists (RAD1, RAD2, and RAD3) were assigned to perform evaluations across all methods. Each radiologist interpreted a subset of 10 randomly assigned images per method, ensuring balanced representation and minimizing bias.

2.4 Outcome Measures and Statistical Analysis

The primary objective of this study was to determine whether AI-generated reports could provide significant time savings in routine, low-risk CXR evaluations, thereby enhancing workflow efficiency. To assess this, the reporting time for each method—manual, template-based, and AI-generated—was recorded from the initial review to the final submission of each report. For each method, the average time per study and the associated standard deviation (SD) were calculated, enabling a comparative analysis that accounts for both central tendency and variability in reporting times.

To evaluate the statistical significance of differences in mean reporting times across the three reporting methods, an analysis of variance (ANOVA) was conducted. This test assessed whether there were overall differences in mean reporting times among the methods. Following a significant ANOVA result, post hoc pairwise comparisons using t-tests were performed to identify specific differences between each pair of methods. These comparisons included measures of mean difference and standard deviation to account for the variability observed in the data. Statistical significance for all tests was defined at a two-sided p-value threshold of 0.05.

3 Results

3.1 Reporting Times for Each Method

The reporting times for each method are presented Table 2, displaying both the total time to report for 10 examinations and the mean time per study (in seconds) for each radiologist. **Manual reporting** requires significantly more time than the other methods, as evidenced by the higher total times recorded by all three radiologists. Specifically, manual reporting averaged approximately 16 minutes and 4 seconds for RAD1, 15 minutes and 10 seconds for RAD2, and 11 minutes and 48 seconds for RAD3 over 10 examinations. The **template-based method** reduced the total reporting time to 5 minutes and 29 seconds for RAD1, 5 minutes and 20 seconds for RAD2, and 8 minutes and 8 seconds for RAD3. This represents an average efficiency improvement of 53.93% compared to manual reporting. The AI-generated reporting method further optimized the reporting process, achieving the lowest average times. **The AI-generated reports** required only 4 minutes and 37 seconds for RAD1, 5 minutes and 19 seconds for RAD2, and 5 minutes and 38 seconds for RAD3 to complete all 10 examinations, corresponding to an average reduction of 62.82% compared to manual reporting. Furthermore, AI-generated reporting showed an additional average efficiency improvement of 15.62% compared to the template-based approach.

When looking at the mean time per individual study, these differences are further highlighted, with manual reporting taking around 96.4 seconds (RAD1), 91 seconds (RAD2), and 70.8 seconds (RAD3) per study. Template-based and AI-generated methods achieved significantly faster times per study, with template-based reporting requiring 32.9 seconds (RAD1), 32 seconds (RAD2), and 48.8 seconds (RAD3), while AI-generated reporting required 27.7 seconds (RAD1), 31.9 seconds (RAD2), and 33.8 seconds (RAD3).

Method	Total Time for 10 Exams (min:s)			Mean Time per Study (s ± SD)		
	RAD 1	RAD 2	RAD 3	RAD 1	RAD 2	RAD 3
Manual Reporting	16:04	15:10	11:48	96.4 ± 5.1	91 ± 4.7	70.8 ± 6.2
Template-Based Reporting	5:29	5:20	8:08	32.9 ± 2.3	32 ± 2.1	48.8 ± 3.5
AI-Generated Reporting	4:37	5:19	5:38	27.7 ± 1.8	31.9 ± 1.6	33.8 ± 2.1

Table 2 – Total reporting time for 10 examinations and mean time per study across methods.

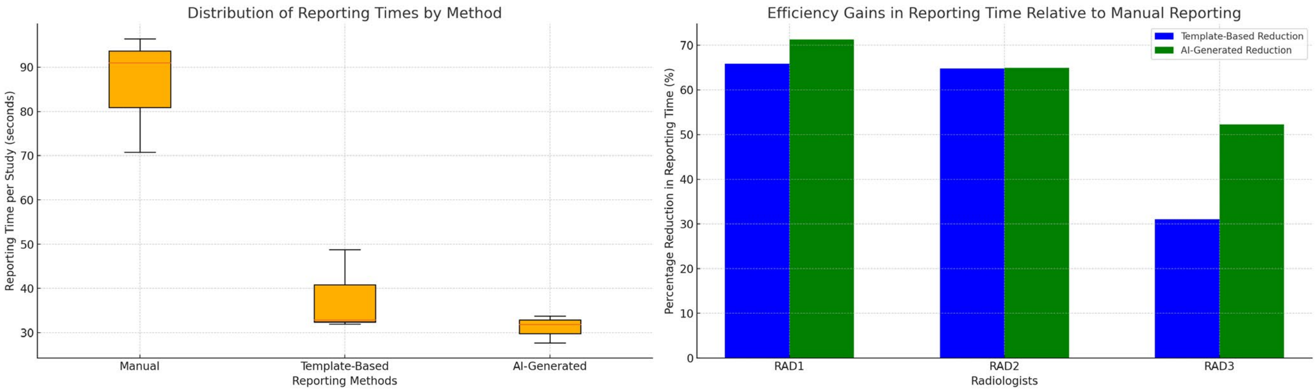


Figure 3 – The box plot (left) illustrates the distribution of reporting times per study for Manual, Template-Based, and AI-Generated methods. The bar chart (right) compares the percentage reduction in reporting times for Template-Based and AI-Generated methods relative to Manual reporting across radiologists.

3.2 ANOVA Results

To determine if there were statistically significant differences in reporting times across the three methods—manual, template-based, and AI-generated—an analysis of variance (ANOVA) was conducted. The ANOVA results, summarized in Table 3, indicate a highly significant effect of the reporting method on the time required to complete reports. The ANOVA indicated a statistically significant effect of reporting method on time ($F(2,6)=28.72$, $p=0.0008$), demonstrating that reporting times differ significantly among methods.

Statistic	Value
F-statistic	28.72
p-value	0.0008

Table 3 – ANOVA results for reporting times across methods.

3.3 Pairwise Comparisons

Following the significant ANOVA result, pairwise *t*-tests were conducted to determine specific differences between each pair of reporting methods: manual, template-based, and AI-generated. These pairwise comparisons allow us to identify which methods differ significantly in terms of reporting time per single study. The results of the *t*-tests, including the *t*-statistics, *p*-values, and significance levels, are summarized in Table 4.

The pairwise comparisons reveal significant reductions in reporting time per single study for both the template-based and AI-generated methods compared to manual reporting. Specifically, the manual vs. template-based comparison yielded a *t*-statistic of 5.06 and a *p*-value of 0.0096, confirming that template-based reporting is significantly faster than manual reporting. Similarly, the manual vs. AI-generated comparison, with a *t*-statistic of 6.87 and a *p*-value of 0.0156, indicates a statistically significant time-saving advantage of AI-generated reporting over the manual method. However, the comparison between template-based and AI-generated methods showed no significant difference, as reflected by a *t*-statistic of 1.18 and a *p*-value of 0.3413.

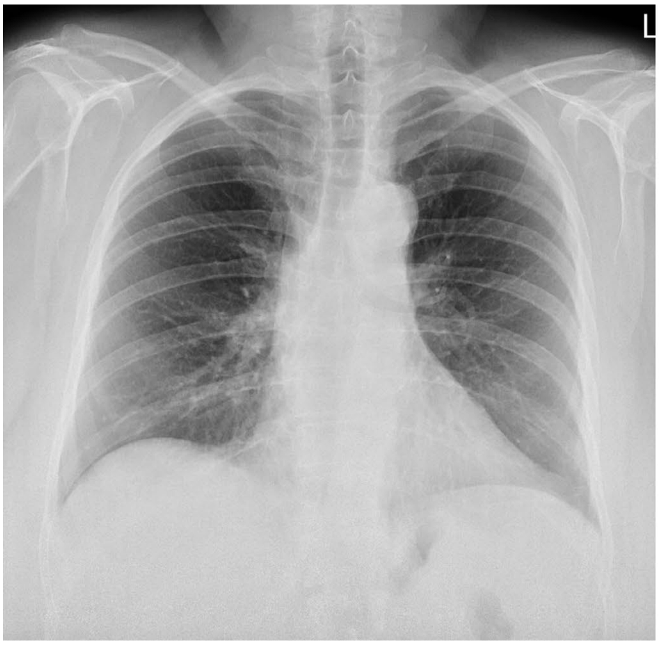
4 Discussion

This study demonstrates that both template-based and AI-generated reporting methods significantly enhance efficiency in low-risk chest X-ray (CXR) reporting compared to traditional manual reporting. The findings highlight how automated and structured methods, particularly AI-generated reporting, can substantially reduce the time radiologists spend on routine cases, thereby potentially improving overall workflow and allowing for reallocation of resources toward complex cases requiring more detailed analysis.

The results showed that manual reporting, which relies on free-text documentation, is the most time-consuming method, with mean times per study of 96.4 seconds (RAD1),

Comparison	*t*-statistic	*p*-value	Mean Difference (± SD)
Manual vs. Template-Based	5.06	0.0096	48.2 ± 16.5
Manual vs. AI-Generated	6.87	0.0156	54.9 ± 13.9
Template-Based vs. AI-Generated	1.18	0.3413	6.8 ± 10.0

Table 4 – Pairwise comparisons of reporting methods (t-tests) with dispersion measures for mean time per study.



(a) Example of the Manual Report in Czech:

> Plicní křídla jsou rozvinutá. Plicní parenchym je bez známek zánětlivých či ložiskových změn. Srdeční stín nerozšířen. Bez známek městnání v malém oběhu. Horní mediastinum bez expanze, bez přesunu. Bránice hladká, klenutá. Zevní CF úhly volné, ostré.

(b) Example of the Template-Based Report in Czech:

> Plíce: Bez známek čerstvých ložiskových nebo infiltrativních změn. Plicní křídla bilaterálně rozvinutá. Plicní kresba přiměřená. Hily v normě.
> Bránice: Bilat. hladké, klenuté. KF úhly bilat. volné.
> Srdce: Srdeční stín nezvětšen.
> Mediastinum: nerozšířeno.

(c) Example of the AI-Generated Report in Czech:

> Plíce: Bez známek atelektázy. Plicní parenchym je bez známek konsolidace. Zevní kostofrenické úhly jsou ostré a volné. Bez detekovaných ložiskových změn v rozsahu plic. Plicní křídla jsou rozvinutá.
> Srdce: Srdeční stín je nezvětšen.
> Ostatní: Bez známek podkožního emfyzému v zobrazeném rozsahu.

Figure 4 – An example of reports generated using three different methods for low-risk chest X-ray evaluations: (a) Manual Report, (b) Template-Based, and (c) AI-Generated report.

91 seconds (RAD2), and 70.8 seconds (RAD3). In contrast, the template-based method, with its structured format, reduced reporting time significantly, with mean times of 32.9 seconds (RAD1), 32 seconds (RAD2), and 48.8 seconds (RAD3). However, AI-generated reporting achieved the highest efficiency gains, offering the lowest mean reporting times per study at 27.7 seconds (RAD1), 31.9 seconds (RAD2), and 33.8 seconds (RAD3).

Statistical analyses confirmed these efficiency gains, with ANOVA showing significant differences in reporting times across the three methods ($F(2,6)$=28.72, p=0.0008). Pairwise comparisons further revealed that both AI-generated and template-based methods offer significant time savings over manual reporting, with mean differences of 48.2 seconds (t=5.06, p=0.0096) and 54.9 seconds (t=6.87, p=0.0156), respectively. However, there was no statistically significant difference in time efficiency between the template-based and AI-generated methods, with a mean difference of 6.8 seconds (t=1.18, p=0.3413). This suggests that while AI provides additional automation, structured templates are also highly effective in expediting low-risk CXR evaluations.

4.1 Limitations

Despite these strengths, the study has several limitations. The single-center design may limit the generalizability of findings to other institutions with differing workflows, technologies, or patient populations. The sample size of 30 CXRs, though sufficient for assessing time efficiency, may not capture the full variability in clinical practice. Furthermore, while the study demonstrated significant time savings, it did not assess diagnostic accuracy or radiologist satisfaction with each method, which are critical factors for successful clinical integration. Future studies should explore these aspects to provide a more comprehensive evaluation of AI-generated reporting.

5 Conclusions

This study concludes that AI-generated reporting for low-risk chest X-rays provides substantial time savings, making it a valuable tool for radiology departments seeking to improve workflow efficiency. Template-based reporting also offers a considerable reduction in reporting time compared to manual methods, underscoring the importance of structured formats in routine reporting. While AI-generated reports do not differ significantly in time efficiency from template-based methods, the automation provided by AI may offer additional benefits in settings with high case volumes or limited radiologist availability.

The integration of AI into routine radiology practice holds promise for streamlining CXR reporting and allowing radiologists to dedicate more time to complex cases. Further research should focus on the long-term impacts of AI integration on radiology workflows and the continuous improvement of AI algorithms to optimize time efficiency across varied clinical settings.

ZVYŠOVÁNÍ EFEKTIVITY PŘI HODNOCENÍ NÍZKORIZIKOVÝCH SKIAGRAMŮ HRUDNÍKU: SROVNÁVACÍ STUDIE MANUÁLNÍCH, ŠABLONOVÝCH A AI GENEROVANÝCH METOD

Abstrakt

Efektivní a přesné hodnocení skiagramů hrudníku (CXR) je zásadní v radiologii, zejména pro rychlou identifikaci nízkorizikových případů a priorizaci složitějších. Tato studie zkoumá časovou efektivitu tří metod hodnocení skiagramů: manuální, šablonové a AI generované, přičemž se zaměřuje na nízkorizikové případy na radiologickém oddělení. Výsledky ukazují, že manuální hodnocení, které vyžaduje volně psanou dokumentaci, trvá výrazně déle než ostatní metody, s průměrnými časy na vyšetření 96,4 sekundy (RAD1), 91 sekund (RAD2) a 70,8 sekundy (RAD3). Strukturovaný přístup založený na šablonách snížil tyto časy na 32,9 sekundy (RAD1), 32 sekund (RAD2) a 48,8 sekundy (RAD3), což představuje průměrné zlepšení efektivity o 53,93 % oproti manuálnímu hodnocení. AI generované hodnocení mělo nejkratší průměrné časy na vyšetření: 27,7 sekundy (RAD1), 31,9 sekundy (RAD2) a 33,8 sekundy (RAD3), což odpovídá průměrnému zkrácení doby o 62,82 % ve srovnání s manuálním hodnocením. Závěrem lze říci, že AI generované hodnocení nabízí významné časové úspory a zároveň si zachovává vysokou přesnost, což naznačuje jeho silný potenciál pro zvýšení efektivity pracovních procesů v radiologii. Studie podporuje integraci AI do rutinního hodnocení CXR, což radiologům umožní soustředit se na složitější případy. Budoucí výzkum by se měl zaměřit na dlouhodobé dopady a další zlepšení AI algoritmů s cílem optimalizovat postupy v radiologii.

Klíčová slova

umělá inteligence, hodnocení skiagramů hrudníku, diagnostické metody hodnocení, radiologie, časová efektivita, efektivita pracovních procesů

References

[1.] Speets, A., Graaf, Y., Hoes, A., Kalmijn, S., Sachs, A., Rutten, M., Gratama, J., Swijndregt, A. & Mali, W. Chest radiography in general practice: indications, diagnostic yield and consequences for patient management. British Journal of General Practice. 56, 574–578 (2006).

[2.] Bansal, T. & Beese, R. Interpreting a chest X-ray. British Journal of Hospital Medicine. 80, C75–C79 (2019).

[3.] Çallı, E., Sogancioglu, E., Ginneken, B., Leeuwen, K. & Murphy, K. Deep learning for chest X-ray analysis: A survey. Medical Image Analysis. 72, 102125 (2021).

[4.] Homayounieh, F., Digumarthy, S., Ebrahimian, S., Rueckel, J., Hoppe, B., Sabel, B., Conjeti, S., Ridder, K., Sistermanns, M., Wang, L. et al. An artificial intelligence–based chest X-ray model on human nodule detection accuracy from a multicenter study. JAMA Network Open. 4, e2141096–e2141096 (2021).

[5.] Kik, S., Gelaw, S., Ruhwald, M., Song, R., Khan, F., Hest, R., Chihota, V., Nhung, N., Esmail, A., Celina Garfin, A. et al. Diagnostic accuracy of chest X-ray interpretation for tuberculosis by three artificial intelligence-based software in a screening use-case: an individual patient meta-analysis of global data. MedRxiv. pp. 2022-01 (2022).

[6.] Kvak, D., Chromcová, A., Ovesná, P., Dandár, J., Biroš, M., Hrubý, R., Dufek, D. & Pajdaković, M. Detecting Pulmonary Lesions in Low-Prevalence Real-World Settings Using Deep Learning. International Conference on Medical Imaging and Computer-Aided Diagnosis. pp. 3–20 (2023).

[7.] Arzamasov, K., Vasilev, Y., Zelenova, M., Pestrenin, L., Busygina, Y., Bobrovskaya, T., Chetverikov, S., Shikhmuradov, D., Pankratov, A., Kirpichev, Y. et al. Independent evaluation of the accuracy of 5 artificial intelligence software for detecting lung nodules on chest X-rays. Quantitative Imaging in Medicine and Surgery. ***14****, 5288 (2024).*

Kontakt

Ing. Marek Řehoř
Fakulta managementu
Vysoká škola ekonomická v Praze
Jarošovská 1117
377 01 Jindřichův Hradec 1
marek.rehor@carebot.com

Šimon Kličník
Department of Radiology
and Nuclear Medicine
University Hospital
Královské Vinohrady
Prague, Czechia

Jakub Dandár

Daniel Kvak
Carebot s.r.o., Prague, Czechia