Original Article

Influencing Factors of Aphasia Severity: An Empirical Study Based on AphasiaBank

Wu Ling Liu Kaipeng

Sichuan International Studies University, Chongqing, China.

Corresponding Author: 2449372741@qq.com

Revised: 01 October 2025 Published: 30 October 2025 Received: 27 August 2025 Accepted: 20 October 2025

Abstract - This study creates a predictive model by methodically analyzing the effects of age, education level, gender, and aphasia type on aphasia severity using data from 232 aphasia patients from the AphasiaBank database. According to the findings, there is no discernible linear correlation between aphasia severity and either age or educational attainment. Nonetheless, decision tree and random forest models show that these two factors, combined with the type of aphasia, can accurately predict severity; the random forest model has an accuracy of 80%. On the Western Aphasia Battery (WAB), female patients scored significantly higher than male patients (p < 0.05), indicating that females have milder aphasia. However, a linear mixed-effects model suggests that sample randomness may have an impact on this finding. This study reveals the multifactorial interplay underlying aphasia severity and provides data-driven support for the development of personalized diagnostic and rehabilitation strategies in clinical practice.

Keywords - Aphasia severity, Influencing factors, Predictive modeling, Random Forest, Western Aphasia Battery.

1. Introduction

Aphasia denotes a core disruption in the capacity to understand or produce language. Focal brain lesions erode previously established language skills, leaving expression, comprehension, reading, and writing compromised. Under the ICF framework, aphasia restricts not only linguistic reception and output but also everyday activities, social engagement, and overall quality of life (Zumbansen, 2014; Zhao & Jiang, 2021). Presentation differs across individuals: global, Broca's, and Wernicke's forms are among the commonly observed variants (Wang et al., 2024).

More than 150 years ago, the first major insight into how language is neutrally grounded came from observing aphasia. Since then, a steady stream of clinical and experimental work on the disorder has continued to clarify how the brain organizes and processes language (Kemmerer, 2015). Stroke is the leading cause of aphasia; roughly 30 % of stroke survivors develop the condition (Saxena, 2017). Traumatic brain injury, brain tumors, and other neurological disorders can also produce aphasia (Zumbansen, 2014). Although the overall area of affected brain tissue often shrinks over time, isolated regions may remain involved. Where the damage occurs largely determines both the type and the severity of aphasia. The most frequent sites are the left cerebral cortex and the basal ganglia (Wang et al., 2024). Aphasia markedly reduces quality of life, disrupting social interaction, provoking depression and anxiety, and restricting everyday activities.

Psychosocial well-being is consistently impaired, and recovery trajectories differ widely across individuals (Griffin, 2020). This study examines how aphasia severity relates to age, education, aphasia subtype, and gender.

2. Influencing Factors

2.1. The Impact of Age on Aphasia

Whether age influences the severity of aphasia is still disputed. Several reports imply that severity increases with age. Li Hongling (2003) observed that aphasic patients frequently present with reduced alertness and cognitive deficits. As age rises, overall cerebral function declines, and the brain's reserve for recruitment shrinks, potentially limiting language recovery and indirectly linking older age to greater aphasia severity.

Younger patients recover language more effectively after stroke, particularly when intervention begins in the acute or subacute window (Brady, 2019). Recovery slows and becomes less complete with advancing age, as the younger brain retains greater plasticity for reorganizing language networks. Among chronic aphasia cases, Harvey (2020) observed that younger individuals tend to sustain gains from high-intensity therapy better, whereas older patients often require extended maintenance sessions. Monnelly (2023) reported that elderly participants tolerate intensive treatment less well, making low-intensity approaches more appropriate.



Younger patients generally retain more cognitive reserve, allowing them to manage language deficits more effectively during recovery (Brady, 2022). Wallentin (2018) reported that non-fluent aphasia cases were on average 52.8 years old, whereas fluent aphasia cases averaged 61.6 years; the 8.8-year gap was significant (t = 4.113, p < 0.001). Earlier studies had already linked advancing age to a higher aphasia risk (Miceli et al., 1981; Pedersen et al., 1995). Engelter et al. (2006) and Tsouli et al. (2009) found a similar pattern: patients aged 61 and older had a significantly higher risk of developing aphasia (t = 4.115, p < 0.001). These authors attribute the rise to declining brain vitality and a higher stroke incidence with age, because stroke often precipitates aphasia, and older age constitutes a key risk factor.

Age's effect on the severity of aphasia is still up for debate, though, as several studies fail to identify a clear correlation. For instance, Yao Jing and colleagues (2013) found no statistically significant change in the distribution of aphasia types across age groups, suggesting that aging does not consistently influence the severity of symptoms. Wang et al. (2013) came to a similar conclusion, stating that age had no effect on the Western Aphasia Battery scores or the classification of aphasia. Gialanella et al. (2011) and Kyrozis et al. (2009) also found no significant correlation between patient age and clinical presentation. However, Croquelois and Bogousslavsky (2011) found that aphasia risk increased after age 65, highlighting the contradictory results in the field.

2.2. Education's Effect on Aphasia

Years of education did not correlate with the presence of speech disorders in aphasics, according to Li et al. (2003). Wang (2024) reached the same conclusion. However, a number of recovery studies have identified education as a factor that can accelerate recovery (He, 1997; Ge & Jiang, 2004). Higher attainment consistently predicts better language outcomes, especially when tasks demand complex processing, because the added cognitive reserve allows patients to absorb and apply rehabilitation strategies more readily. Moreover, better-educated patients tend to show stronger self-management and self-motivation, attitudes that help sustain engagement throughout therapy (Brady, 2022). Zhou Li and Ye Jing (2013) likewise found no link between years of schooling and post-stroke language recovery. Wallentin (2018) observed that educational level does not predict how often aphasia occurs, how severe it is, or which type appears. Cameron et al. (1981) reported that individuals with little formal education are less prone to aphasia after left-hemisphere injury. Because language representation in unschooled brains tends to be more bilateral, extensive left-sided damage does not always erase verbal ability; moreover, their smaller vocabularies can mask subtle linguistic deficits.

2.3. The Impact of Aphasia Types on Aphasia

Besides age, gender, and education, lesion site, stroke subtype, and cognitive impairment also shape aphasia recovery (Wang et al., 2024; Xu et al., 2023). cognitive

impairment, etc., also have a significant impact on recovery from aphasia (Wang et al., 2024; Xu et al., 2023). The location of the lesion and the type of stroke directly affect the language center of the brain, thereby influencing the recovery of language function (Wang et al., 2024). The presence of cognitive impairment may further limit the rehabilitation potential of patients and requires special attention in treatment (Xu et al., 2023).

Zhang Yumei et al. (2005) noted that aphasia categories do not always map onto the classical language centers. They found that lesions outside these zones can still produce aphasia, yet when the centers themselves are damaged, the deficit is usually more severe and clinically striking. Except for female patients with cerebral hemorrhage, motor, complete, and named aphasia were the most common. In addition, some scholars have found that age and stroke type have an impact on the type of aphasia (Yao Jingfan & Zhang Yumei, 2015).

2.4. Gender Influence on Aphasia

Most studies agree that gender has little bearing on how severe aphasia is overall, though it can shape particular symptoms. Chen Ying and Li Yansheng (2009) found that the likelihood of post-stroke aphasia shifts with both sex and age: men outnumbered women in the 51-70-year band, whereas women predominated after 70. Despite this, the relative frequency of aphasia subtypes stayed the same for both sexes, with global and Broca's aphasia the most common. DeRenzi et al. (1980) reached a similar conclusion, detecting no sexlinked contrast. Hier et al. (1994), however, observed a higher risk in women, giving a female-to-male ratio of 1.35:1 in Western samples. Yao Jing and colleagues (2013) again recorded more cases among Chinese men, yet the subtype profile was almost identical for the two sexes, with Broca's aphasia leading in both. Taken together, the evidence implies that sex exerts only a minor influence on how profound the language impairment becomes.

Zhang Yumei (2005) found that men had a consistently higher rate of post-stroke aphasia in every age band, and the gap reached statistical significance; among the middle-aged and elderly, it widened to 82.13 %. After the age of 60, however, women's incidence climbed steadily to 59.17 %, erasing or even reversing the male excess. One reason for the sex-specific pattern may lie in how language is lateralized: females more often engage both hemispheres. During letter fluency tasks, for example, men recruit left-hemisphere regions almost exclusively, whereas women co-activate homologous right-sided areas. This bilateral layout can leave verbal ability partly intact when the left hemisphere is damaged. Women between the ages of 50 and 60 have a lower rate of aphasia, which is associated with a lower risk of stroke. This effect is commonly attributed to the neuro- and vascular protective properties of estrogen. When combined, the findings suggest that there are gender differences in the

incidence and recovery of aphasia, with women typically recovering specific language skills more quickly. This advantage is probably due to a greater degree of lateralization of their language circuitry (Brady, 2022).

In aphasia, age, gender, and education all play a role in determining the kind and severity of language loss as well as the extent of therapy and the rate at which improvements last.

Clinicians can improve everyday communication and long-term well-being by mapping these intersecting influences and creating interventions that are tailored to the individual rather than the chart (Brady, 2020; Wang et al., 2024; Xu et al., 2023).

3. Study Design

3.1. Research Questions

Three questions are formulated in this study after a thorough review of pertinent literature. First, is there a significant correlation between age, education, aphasia type, and aphasia severity? Second, is there a statistically significant relationship between severity and gender? Third, is it possible to combine aphasia type, age, and education to create a machine learning model that accurately predicts severity?

3.2. Research Subjects and Data Sources

This study drew on records from AphasiaBank, a shared repository that archives standardized clinical data contributed by people with aphasia. All datasets follow a common collection protocol and include speech samples, picture descriptions, narrative productions, discourse excerpts, and scores from instruments such as the Western Aphasia Battery-Revised (WAB-R) (Sharma, 2019, p. 808).

Data from AphasiaBank were analyzed to explore how age, education, aphasia type, and gender might relate to severity. The study first filtered each participant's demographic record and WAB score, removed entries with missing details, and then aligned the remaining data to yield 232 complete cases.

3.3. Research Tools and Methods

The analysis relied on R statistical software and the AphasiaBank corpus. WAB scores were first recoded into mild, moderate, or severe levels according to Fromm et al. (2017), while age and education were binned with the criteria of Wang Yuyuan (2013). Variance homogeneity was met, yet normality was rejected.

Correlations among age, education, aphasia type, and severity were screened before age, education, and type were supplied to a decision tree and a random forest. Gender differences in WAB scores were finally examined with a rank-sum test and a linear mixed-effects model, chosen to accommodate the unequal variances.

4. Results and Discussion

4.1. Effects of Age, Educational Level, and Aphasia Type on Aphasia Severity

Correlation tests showed that neither age nor education was significantly linked to WAB scores (r = -0.11 and 0.16, respectively), possibly because the relationships are not linear. To explore this, we fitted decision-tree and random-forest models using age, education, aphasia type, and aphasia severity; the outcome is plotted in Figure 1. Aphasia type emerged as the strongest predictor of severity. Single trees are unstable, so we ensemble them into a random forest. As Figure 2 illustrates, the combined model reaches an out-of-bag error of 21.98 % on the AphasiaBank corpus, well below the baseline, confirming that the variables jointly carry predictive value for aphasia severity.

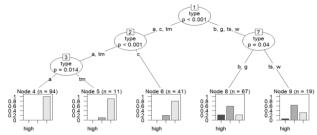


Fig. 1 Decision tree for Age, Educational Level, and Aphasia Type on Aphasia Severity

Note: a: Nominal aphasia b: Broca aphasia c: Conductive aphasia g: Global aphasia tm: Motor aphasia ts: Sensory aphasia w: Wilnick aphasia

4.2. Gender Influence on the Severity of Aphasia

A Kolmogorov–Smirnov test on the WAB scores returned D = 0.093406 and p < 0.05, so the distribution deviated from normality. Levene's test gave F = 4.99, p = 0.02, pointing to unequal variances. We used the Wilcoxon rank-sum test after both assumptions were broken, and the results showed that there was a significant difference in the severity of aphasia between the sexes (W = 6675.5, p < 0.05).

We fitted a linear mixed-effects model to measure the gender gap. Each patient contributed a single observation, so the experiment was entered as a random effect, and gender was treated as a fixed effect. Females scored 9.344 points higher on the WAB than males, yet the large intercept signals pronounced random-effect variance, casting doubt on the estimate's stability. This suggests that the gender association could be distorted by unmeasured confounders, an issue that later work should address.

4.3. Discussion

Across the three research questions, age and education showed no measurable influence on aphasia severity. The outcome aligns with Yao Jing et al. (2013) and Wang Yuyuan et al. (2013), yet diverges from Li Hongling (2003), Brady (2019, 2021), and Wallentin (2018). One reason the age effect vanished here may be the uneven burden of comorbidities among older participants, masking any age-linked gradient. Another is that contemporary rehabilitation appears to help younger and older patients equally, diluting age as an isolated factor. Consequently, age seems less central than interindividual variation in neural reserve and rehabilitation response.

Gender exerted a statistically significant influence on aphasia severity in this study, echoing the findings reported by Chen Ying and Li Yansheng (2009), Yao Jing et al. (2013), Zhang Yumei (2005), and Hier et al. (1994). Because homogeneity of variance was violated, the outcome demands cautious interpretation and calls for replication in larger samples. Last but not least, a model that combined age, education, and aphasia type showed a reasonably accurate

prediction of aphasia severity, indicating that these three variables work well together as predictors.

5. Conclusion

Using the AphasiaBank dataset for further analysis, this study examined age, education, gender, and aphasia subtype as possible predictors of aphasia severity. While gender influenced outcomes—women scored higher than men on the WAB, though the difference may have been mitigated by unmeasured factors—age and education did not correlate with severity. To identify this gender effect, larger samples are needed. On the other hand, the combination of aphasia type, age, and education produced a useful severity prediction. A small sample size and inadequate control for potential confounding variables outside of those studied are two of the study's limitations. To gain a better understanding of the factors that influence the results of aphasia rehabilitation, future studies should increase the sample size, use multivariate statistical techniques, and include a wider range of potential influencing factors.

References

- [1] The RELEASE Collaborators et al., "Precision Rehabilitation for Aphasia by Patient Age, Sex, Aphasia Severity, and Time Since Stroke? A Prespecified, Systematic Review-based, Individual Participant Data, Network, Subgroup Meta-analysis," *International Journal of Stroke*, vol. 17, no. 10, pp. 1067–1077, 2022. [CrossRef] [Google Scholar] [Publisher Link]
- [2] RELEASE Collaboration, "Communicating Simply, But not too Simply: Reporting of Participants and Speech and Language Interventions for Aphasia After Stroke," *International Journal of Speech-Language Pathology*, vol. 22, pp. 302–312, 2020. [CrossRef] [Google Scholar] [Publisher Link]
- [3] Marian C. Brady et al., "RELEASE: A Protocol for a Systematic Review Based, Individual Participant Data, Meta- and Network Meta-analysis, of Complex Speech-language Therapy Interventions for Stroke-related Aphasia," *Aphasiology*, vol. 34, no. 2, pp. 137–157, 2019. [CrossRef] [Google Scholar] [Publisher Link]
- [4] Ralph. F. Cameron, Robert D. Currier, and Armin F. Haerer, "Aphasia and Literacy," *British Journal of Disorders of Communication*, vol. 6, no. 2, pp. 161–163, 1971. [Google Scholar] [Publisher Link]
- [5] Alexandre Croquelois, and Julien Bogousslavsky, "Stroke Aphasia: 1,500 Consecutive Cases," *Cerebrovascular Diseases*, vol. 31, no. 4, pp. 392–399, 2011. [CrossRef] [Google Scholar] [Publisher Link]
- [6] E. De Renzi, P. Faglioni, and P. Ferrari, "The Influence of Sex and Age on the Incidence and Type of Aphasia," *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, vol. 16, no. 4, pp. 627–630, 1980. [CrossRef] [Google Scholar] [Publisher Link]
- [7] Stefan. T. Engelter et al., "Epidemiology of Aphasia Attributable to First Ischemic Stroke: Incidence, Severity, Fluency, Etiology, and Thrombolysis," *Stroke*, vol. 37, no. 6, pp. 1379–1384, 2006. [CrossRef] [Google Scholar] [Publisher Link]
- [8] Davida Fromm et al., "Discourse Characteristics in Aphasia Beyond the Western Aphasia Battery Cutoff," *American Journal of Speech-Language Pathology*, vol. 26, no. 3, pp. 762–768, 2017. [CrossRef] [Google Scholar] [Publisher Link]
- [9] Bernardo Gialanella, "Aphasia Assessment and Functional Outcome Prediction in Patients with Aphasia After Stroke," *Journal of Neurology*, vol. 258, pp. 343–349, 2011. [CrossRef] [Google Scholar] [Publisher Link]
- [10] Jenna. R. Griffin-Musick et al., "The Impact of a University-based Intensive Comprehensive Aphasia Program (ICAP) on Psychosocial Well-being in Stroke Survivors with Aphasia," *Aphasiology*, vol. 35, no. 12, pp. 1363–1389, 2020. [CrossRef] [Google Scholar] [Publisher Link]
- [11] Sam R. Harvey et al., "Treatment Dose in Post-stroke Aphasia: A Systematic Scoping Review," *Neuropsychological Rehabilitation*, vol. 31, no. 10, pp. 1629–1660, 2020. [CrossRef] [Google Scholar] [Publisher Link]
- [12] Daniel B. Hier et al., "Gender and Aphasia in the Stroke Data Bank," *Brain and Language*, vol. 47, no. 1, pp. 155–167, 1994. [CrossRef] [Google Scholar] [Publisher Link]
- [13] David Kemmerer, Cognitive Neuroscience of Language, New York, NY, USA: Psychology Press, 2022. [CrossRef] [Google Scholar] [Publisher Link]

- [14] A. Kyrozis et al., "Incidence and Predictors of Post-stroke Aphasia: The Arcadia Stroke Registry," *European Journal of Neurology*, vol. 16, no. 6, pp. 733–739, 2009. [CrossRef] [Google Scholar] [Publisher Link]
- [15] Gabriele Miceli et al., "Influence of Age, Sex, Literacy and Pathologic Lesion on Incidence, Severity and Type of Aphasia," *Acta Neurologica Scandinavica*, 1981. [CrossRef] [Google Scholar] [Publisher Link]
- [16] Katie Monnelly et al., "Intensive and Comprehensive Aphasia Therapy—A Survey of the Definitions, Practices and Views of Speech and Language Therapists in the United Kingdom," *International Journal of Language & Communication Disorders*, vol. 58, no. 6, pp. 2077-2102, 2023. [CrossRef] [Google Scholar] [Publisher Link]
- [17] Palle Moller Pedersen et al., "Aphasia in Acute Stroke: Incidence, Determinants, and Recovery," *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, vol. 38, no. 4, pp. 659–666, 1995. [CrossRef] [Google Scholar] [Publisher Link]
- [18] Sadhvi Saxena, and Argye Hillis, "An Update on Medications and Noninvasive Brain Stimulation to Augment Language Rehabilitation in Post-stroke Aphasia," *Expert Review of Neurotherapeutics*, vol. 17, no. 11, pp. 1091–1107, 2017. [CrossRef] [Google Scholar] [Publisher Link]
- [19] Saryu Sharma et al., "Gender Differences in Aphasia Outcomes: Evidence from the AphasiaBank," *International Journal of Language & Communication Disorders*, vol. 54, no. 5, pp. 806-813, 2019. [CrossRef] [Google Scholar] [Publisher Link]
- [20] S. Tsouli et al., "Significance of Aphasia after First-ever Acute Stroke: Impact on Early and Late Outcomes," *Neuroepidemiology*, vol. 33, no. 2, pp. 96–102, 2009. [CrossRef] [Google Scholar] [Publisher Link]
- [21] Mikkel Wallentin, "Sex Differences in Post-stroke Aphasia Rates are Caused by Age. A Meta-analysis and Database Query," *PloS One*, vol. 13, no. 12, 2018. [CrossRef] [Google Scholar] [Publisher Link]
- [22] Anna Zumbansen, and Alexander Thiel, "Recent Advances in the Treatment of Post-stroke Aphasia," *Neural Regeneration Research*, vol. 9, no. 7, pp. 703–706, 2014. [CrossRef] [Google Scholar] [Publisher Link]
- [23] Chen Ying, and Li Yan-Sheng, "The Effect of Gender on Post-stroke Aphasia," *Journal of Shanghai Jiao Tong University (Medical Science)*, vol. 29, no. 8, pp. 978–981, 2009. [Publisher Link]
- [24] Ge Hong, and Jiang Jin, "A Survey on the Relationship Between Language Rehabilitation and Social Factors in Patients with Aphasia," *Journal of Soochow University (Medical Edition)*, vol. 24, no. 6, pp. 829–831, 2004. [Publisher Link]
- [25] He Xiaoying, "The Relationship Between Treatment of Aphasia and Various Factors," *Journal of Practical Cardiovascular and Cerebrovascular Diseases*, vol. 1, pp. 42-43, 1997. [Publisher Link]
- [26] Li Hongling et al., "The Occurrence and Influencing Factors of Speech Dysfunction in Stroke Patients," *Chinese Journal of Rehabilitation Medicine*, no. 8, pp. 473-475, 2003. [Publisher Link]
- [27] Wang Yuyuan, Sun Changhui, and Bai Yulong, "A Retrospective Analysis of Factors Influencing Types and Severity of Aphasia in Patients with Post-stroke Aphasia," *Chinese Journal of Rehabilitation Medicine*, vol. 6, pp. 323-328, 2024. [Publisher Link]
- [28] Xu Ling, and Wang Tao, "Research Progress on Screening Tools for Post-stroke Aphasia with Cognitive Impairment," *Nursing Research*, no. 1, pp. 82–87, 2023. [Publisher Link]
- [29] Yao Jingfan, and Zhang Yumei, "Research on Factors Related to the Types of Aphasia after Stroke," *Proceedings of the 12th Beijing International Rehabilitation Forum*, 2022. [Publisher Link]
- [30] J.F. Yao et al., Research on the Associated Factors of Types of Post-stroke Aphasia, Compilation of Papers from the 18th National Conference on Neurology of the Chinese Medical Association, 2015. [Online]. Available: https://kns.cnki.net/kcms2/article/abstract?v=IPzr95zWmwQPNixDfGnDnk1r_1543-NuP6FC8UgR4iwDbt7PhQF9_U9pY_eRQkO1BFeuO4Bwwcmgh6yP_KwaFWTzoAFXduStFNkq96fXfsqtMlTpfnWILu8rjinIVdugzApcGgu PRTO5VguV K RtpQ7dkp4GptypVbeqOxi2UOBgRgVDGw6Q==&uniplatform=NZKP
- [31] Y.M. Zhang et al., "Analysis of Factors Influencing the Type of Aphasia in Stroke Units," *Proceedings of the 9th National Conference on Cerebrovascular Disease Rehabilitation*, pp. 196–197, 2005. [Online]. Available: https://kns.cnki.net/kcms2/article/abstract?v=IPzr95zWmwQuqANNBoGDB_ExQbGgbXKDQvKNRBxV4eZv_X249AynjO-kQzB4ER3HAJpxgE-cgxF6IbMYLqJiDhTgZQVy4ury-eKOJPHDIkG4EuXx1u-AuacBieclAlP0wY-dC0Onxla6DApe3aTv86WRzUsef9Vu7fjoqVLRnF8=&uniplatform=NZKPT
- [32] Zhao Nisha, and Jiang Meng, "Research on Language Barriers among Chinese-Speaking Aphasics in China Over the Past Thirty Years," *Journal of Audiology and Speech Disorders*, vol. 29, no. 4, pp. 464–467, 2021. [Publisher Link]
- [33] Liye Zhou, and Jing, "Analysis of Prognosis and Influencing Factors of Aphasia After Stroke," *Chinese Medical Guide*, no. 19, pp. 22–24, 2013. [Publisher Link]