Shruti: 22 small offline speech-to-text models, one per language

UDGAM LABS VERSION 1.0

Abstract

Shruti is a family of 22 speech-to-text models. Each model is one small file for one language plus English: 29.5 MB for Shruti Hindi and 29.7 MB for the other 21, about 56M parameters each. Each writes its language in that language's script and tries to keep English words in lowercase English letters, the way many people type. All of them run offline in Cactus Compute's open-source needle engine, on a phone, a laptop or in a browser. All are full fine-tunes built on Cactus Whistle (Apache-2.0).

We ran every model in this report ourselves, on the same clips with the same scorer for each language. The results differ a lot from language to language, so this report does not give a family headline for accuracy.

- **What holds across the family.** One file, offline. On a five-year-old phone with 6 GB of RAM, 8.6x to 12.6x faster than real time in 64 MB of memory on the 21 newer models; Shruti Hindi was measured differently, at 6.3x on a 60-second recording in 80 MB. No words invented on 100 silent clips, for all 22.
- **Where it leads.** On conversations scored with English words in English letters, several models make fewer errors than rivals many times their size. Section 6 gives each result with its test and its size. Those leads depend on that way of scoring.
- **Where it falls short.** On read speech, and against references that write English words in the language's own script, larger models are ahead on most languages, often by a wide margin. On some conversation tests more than half of the words are wrong. Five languages have no claim against a serious rival, and five have only a weak one.
- **ONNX packs.** All 22 also come as ONNX packs for ONNX Runtime, the browser and sherpa-onnx. Tested against the engine under six written rules, none passed them all: 13 are mixed and 9 failed. Section 10 gives each language's result. Every score above is the engine's.

1. The family

Each Shruti model covers one language and English, including speech that mixes the two. Hindi in Devanagari with English in Latin looks like this: "मेरा order कब आएगा".

Each model is its own file. A model does not transcribe the other languages in the family: Shruti Marathi does not do Hindi, for example. Pick the model for the language you need.

LANGUAGE	WRITES THE LANGUAGE IN	FILE	PARAMETERS	VOCABULARY PIECES ADDED
Hindi	Devanagari	`shruti.cact`, 29.5 MB	55M	714
Marathi	Devanagari	`shruti-marathi.cact`, 29.7 MB	55,722,313	400
Tamil	Tamil script	`shruti-tamil.cact`, 29.7 MB	55,875,913	700
Bengali	Bengali script	`shruti-bengali.cact`, 29.7 MB	55,875,913	700
Telugu	Telugu script	`shruti-telugu.cact`, 29.7 MB	55,875,913	700
Gujarati	Gujarati script	`shruti-gujarati.cact`, 29.7 MB	55,875,913	700
Odia	Odia script	`shruti-odia.cact`, 29.7 MB	55,875,913	700
Kannada	Kannada script	`shruti-kannada.cact`, 29.7 MB	55,875,913	700
Urdu	Perso-Arabic	`shruti-urdu.cact`, 29.7 MB	55,875,913	700
Malayalam	Malayalam script	`shruti-malayalam.cact`, 29.7 MB	55,875,913	700
Assamese	Assamese script (Bengali-Assamese block)	`shruti-assamese.cact`, 29.7 MB	55,875,913	700
Sanskrit	Devanagari	`shruti-sanskrit.cact`, 29.7 MB	55,722,313	400
Punjabi	Gurmukhi	`shruti-punjabi.cact`, 29.7 MB	55,875,913	700
Bodo	Devanagari	`shruti-bodo.cact`, 29.7 MB	55,722,313	400
Dogri	Devanagari	`shruti-dogri.cact`, 29.7 MB	55,722,313	400
Kashmiri	Perso-Arabic	`shruti-kashmiri.cact`, 29.7 MB	55,875,913	700
Konkani	Devanagari only, not Romi or Kannada script	`shruti-konkani.cact`, 29.7 MB	55,722,313	400

LANGUAGE	WRITES THE LANGUAGE IN	FILE	PARAMETERS	VOCABULARY PIECES ADDED
Maithili	Devanagari	`shruti-maithili.cact`, 29.7 MB	55,722,313	400
Manipuri	Meetei Mayek only, not Bengali script	`shruti-manipuri.cact`, 29.7 MB	55,875,913	700
Nepali	Devanagari	`shruti-nepali.cact`, 29.7 MB	55,722,313	400
Santali	Ol Chiki only	`shruti-santali.cact`, 29.7 MB	55,875,913	700
Sindhi	Devanagari only, not Perso-Arabic	`shruti-sindhi.cact`, 29.7 MB	55,722,313	400

Shruti (श्रुति) is Sanskrit for "that which is heard".

---

2. How it is built

2.1 Lineage

- **Cactus Whistle** by Cactus Compute (Apache-2.0) is the base model.
- **Shruti Hindi** is a full fine-tune of Whistle: all its weights were trained, and 714 Devanagari vocabulary pieces were added.
- **Each other language** is a full fine-tune of Shruti Hindi's weights. We added vocabulary pieces for the language (the table above) and reused Shruti Hindi's Hindi language tag for it.

Every model is therefore a modified version of Whistle, and each `NOTICE` file says so. The file is exported with 4-bit weights and 8-bit activations, and the results in this report were measured on the exported file, in the engine.

2.2 Architecture

The architecture is Whistle's, described in full in the Shruti Hindi report:

- an encoder-decoder transformer with 8 encoder and 8 decoder layers of width 512;
- 80-band log-mel input from 16 kHz audio, with a convolutional stem that gives one frame every 80 ms;
- output of a language tag and then text, one word piece at a time;

- 16 kHz audio of any length as input; one pass covers up to 30 seconds, and longer audio is processed in 30-second windows;
- no punctuation in the output, and numbers written as words.

2.3 Training data, at a high level

Each model trains on public speech datasets for its language plus English from Shruti Hindi's training data. Section 9 names the sources for each language and their licences. This report does not describe data amounts, proportions, training settings, compute or cost.

---

3. How to run it

Every model runs in the Cactus needle engine (Apache-2.0). No server and no internet connection are needed once the file is on the device.

Python:

```
pip install cactus-needle
hf download udgamlabs/shruti-bodo shruti-bodo.cact config.json --local-dir .
```

```
from needle import Whistle

model = Whistle("shruti-bodo.cact")
print(model.transcribe("clip.wav"))                    # the language is chosen
automatically
print(model.transcribe("english.wav", language="en"))  # only for speech that
is entirely English
```

The Python package sends anonymous usage counts by default; set `NEEDLE_TELEMETRY=0` to turn that off. The command-line engine and the WebAssembly build make no network calls.

With our examples (github.com/udgamlabs/shruti):

```
python examples/transcribe.py clip.wav --model marathi
python examples/transcribe.py clip.wav --model nepali --language ne
SHRUTI_MODEL=marathi ./examples/transcribe.sh clip.wav
```

In a browser: `examples/browser/` is one page with no build step. It runs the engine's WebAssembly build in a Web Worker and transcribes on the device. The page takes `?model=<language>`. A working page is at udgamlabs.com/models/shruti.

In an app: link the engine's C library and ship the model file with the app, or use an ONNX pack. Section 10 has the steps for both, and each pack's test result.

Language codes as the site's picker sends them: Hindi `hi`, Marathi `mr`, Tamil `ta`, Bengali `bn`, Telugu `te`, Gujarati `gu`, Odia `or`, Kannada `kn`, Urdu `ur`, Malayalam `ml`, Assamese `as`, Sanskrit `sa`, Punjabi `pa`, Bodo `brx`, Dogri `doi`, Kashmiri `ks`, Konkani `kok`, Maithili `mai`, Manipuri `mni`, Nepali `ne`, Santali `sat`, Sindhi `sd`. Set the language for Nepali when you know it is Nepali (section 8.2).

Where we measured it. For the 21 newer models: a Mac (native), an Android phone (native engine) and that phone's Chrome (WebAssembly). Shruti Hindi was also measured on Linux. The other 21 have not been measured on Linux, Windows or iOS.

---

4. How we evaluated it

4.1 One harness per language

For each language, every model ran on the same clips and went through the same text normaliser and scorer: Unicode NFC, lowercase, punctuation removed, then word error rate (WER, %, lower is better) as written. A WER above 100 means the model added words. Shruti models ran in the on-device engine on a Mac; rivals ran through the same harness. File sizes are the file as published by its makers (whisper.cpp q8_0 for Whisper tiny, base and small). Measurements are dated in each section; the dates run from 18 October to 6 November 2026.

Shruti's results use the engine's own language choice, unless a section says otherwise. Shruti Hindi's main table ran with the language set.

4.2 Test sets

SET	WHAT IT IS	HOW TO READ IT
FLEURS test	Read Wikipedia sentences	We did not train on it, so it is an independent read test. The exception is Shruti Hindi, which trained on FLEURS's training split, so FLEURS is closer to its training data. It appears only for the languages whose section 6 table has a FLEURS column
Kathbath test-unknown	Read sentences; speakers not in training	Its speakers are unseen, but Kathbath's training split is in our training data, and very likely the rivals' too. Shruti Marathi used Kathbath's valid split instead, with its speakers kept out of training
DiarBench	Real conversation, single-speaker turns; held out from every model's training	The independent check, and the lead set when a claim is challenged
IndicVoices valid	IndicVoices' own held-out split: read, extempore and conversation	Speakers kept out of training and sentences blocked, but **the same source as our training data, and very likely IndicConformer's, so it flatters both.** It is not a read-speech test in the FLEURS sense
Monsoon, Lahaja, MUCS, Svarah	Used for Shruti Hindi; section 6.1	Hindi, accented Hindi, Hindi-English code-mix, Indian English
LibriSpeech test-clean	Read US English, 2,620 clips	Run with English set
Silence	100 silent clips; we count the words a model writes	Silence only, not noise or music

For the 21 newer models, sets hold up to 200 clips per language.

DiarBench is not one uniform set. For Hindi, Marathi, Tamil, Bengali, Telugu, Gujarati, Odia, Kannada, Urdu, Malayalam, Assamese and Punjabi, it holds only turns with at least one English word. For Sanskrit, Bodo, Dogri, Kashmiri, Konkani, Maithili, Manipuri, Nepali, Santali and Sindhi, it holds every usable turn, with or without English. So "conversations that mix in English" means different things across the family, and the sizes differ a lot:

LANGUAGE	TURNS	CONVERSATIONS	ENGLISH WORDS	TURNS WITH ENGLISH	READS AS
Hindi	200	57	1,553	200	Big enough for its claim
Marathi	200	42	805	200	Big enough for its claim
Tamil	200	50	987	200	Big enough for its claim
Bengali	200	64	770	200	Big enough for its claim
Telugu	200	50	1,016	200	Big enough for its claim
Gujarati	200	43	907	200	Big enough for its claim
Odia	196	37	1,459	196	Big enough for its claim
Kannada	200	49	1,091	200	Big enough for its claim
Urdu	74	25	251	74	Small
Malayalam	200	34	1,038	200	Big enough for its claim
Assamese	84	19	375	84	Small
Sanskrit	55	21	68	15	Small; too few English words
Punjabi	200	44	659	200	Big enough for its claim
Bodo	112	25	409	97	Small
Dogri	24	17	40	16	Too small to carry a claim
Kashmiri	32	14	0	0	Too small; holds no English
Konkani	90	23	278	71	Small
Maithili	54	15	109	30	Small; too few English words
Manipuri	44	20	243	39	Too small to carry a claim
Nepali	45	19	118	34	Too small; too few English words
Santali	99	28	63	34	Small; too few English words
Sindhi	27	17	51	12	Too small; too few English words

The floors behind that last column are 50 turns and 200 English words. A small set carries a claim only with its turn count stated, and only when the intervals by clip and by conversation both exclude 0.

4.3 What each reference means

Conversations are scored against two references, and they give different results:

- **English in Latin.** The reference writes English words in English letters, as DiarBench's source glosses them. An English word counts as right only if it is written in English letters. This is the convention Shruti is built for, and it is the one where its leads appear.
- **Native.** The reference writes English words in the language's own script. A model that keeps "cinema" in English letters is marked wrong here, even when it heard the word correctly, and rivals that write only the native script are not penalised for it.

Both are shown for every language, because a lead on one can reverse on the other. IndicConformer, for example, writes almost no Latin at all. DiarBench also glosses some local names in Latin, so a correct native spelling of a name can count as an error.

English words in English letters is the share of a reference's English words that a model wrote in Latin letters.

We leave one thing out of the tables: IndicVoices valid with a Latin reference. That reference carries event tags and bracketed glosses, which the scorer counts as extra words, so no claim may use it. The exception is Shruti Marathi, whose older result is shown and marked in section 6.2.

4.4 Is a difference real?

For most languages we report paired 95% bootstrap intervals of WER (Shruti minus the rival, in points, over the same clips, 1,000 resamples): negative means fewer errors for Shruti, and an interval that includes 0 shows no lead either way. On DiarBench we ran up to three checks. For Hindi, Marathi, Tamil, Bengali and Telugu they were run on 7 November 2026 from saved transcripts, with no model re-run:

- **By clip**, the basic check.
- **By conversation**, resampling whole conversations, because turns from one conversation are not independent.
- **Script-blind**, on a romanised view that ignores which script a word is in. If a lead vanishes here, it is only about writing English in English letters. This check does not work for Perso-Arabic, Meetei Mayek or Ol Chiki, so Urdu, Kashmiri, Manipuri and Santali do not have it.

4.5 Rivals

For each language we measured the open models we could run, listed in its section. Many other models exist that we have not measured, and section 8.7 names the ones most likely to matter. Claims always mean "the models we measured". Whisper's error rates above 100% mostly come from runaway or wrong-script output, so they say little about Whisper at its best.

4.6 The verdicts

Each newer language carries a verdict from its facts file. It is our own rule of thumb:

- CLAIM: the paired interval excludes 0 against the language's main rival on DiarBench, by clip and by conversation, with enough turns and English words.
- WEAK CLAIM: true but limited: too few turns, it fails the conversation or script-blind check, it appears only on IndicVoices valid, or it is only "no clear difference".
- NO CLAIM: no result against a serious rival.

Shruti Hindi, Marathi, Tamil, Bengali, Telugu and Gujarati were released before this labelling. Their sections state the claims their facts files allow, with no label, and their DiarBench checks.

5. The family at a glance

LANGUAGE	VERDICT	WHAT THE EVIDENCE SUPPORTS, IN ONE LINE	WHERE IT IS BEHIND, IN ONE LINE
Hindi	No label	Fewer errors than Whisper large-v3 on three Hindi sets, in a file about 100x smaller; lowest Hinglish error of the models measured on MUCS	Koyal and IndicConformer Hindi on pure Hindi; Whisper large-v3 on English
Marathi	No label	The most accurate open Marathi model under 100 MB that we measured (2 Nov 2026)	IndicConformer Marathi on read Marathi and on native references; Hima-ASR's mixed build on FLEURS, Kathbath and DiarBench
Tamil	No label	The most accurate open Tamil model under 100 MB that we measured (2 Nov 2026)	IndicConformer Tamil and SraVaani-1.0 on read Tamil and on the native reference
Bengali	No label	Fewer errors than Vosk Bengali and IndicConformer Bengali on mixed conversations; the IndicConformer lead needs English in English letters	Both on read Bengali and the native reference; SraVaani-1.0 on every set
Telugu	No label	Fewer errors than IndicConformer Telugu and Whisper Telugu fine-tunes on mixed conversations; level with SraVaani-1.0	Most larger rivals on read Telugu and the native reference
Gujarati	No label	Fewer errors than every model we measured on mixed conversations	IndicConformer Gujarati and SraVaani-1.0 on read Gujarati; SraVaani-1.0 on the native reference
Odia	CLAIM	Fewer errors than every model we measured on mixed conversations	IndicConformer Odia and SraVaani-1.0 on read Odia and the native reference
Kannada	CLAIM	Fewer errors than IndicConformer Kannada and Whisper Kannada fine-tunes on mixed conversations	Over half of words wrong on that test; most rivals ahead on read Kannada
Urdu	CLAIM	About 21% fewer errors than IndicConformer Urdu on 74 mixed conversation turns	That lead rests on English letters; IndicConformer Urdu ahead elsewhere
Malayalam	WEAK CLAIM	One small file for Malayalam and English; no comparison in its headline	Over half of words wrong on conversations; IndicConformer Malayalam and SraVaani-1.0 ahead on read Malayalam
Assamese	CLAIM	Fewer errors than IndicConformer-600M and SraVaani-1.0 on 84 mixed conversation turns	Both ahead on read Assamese and the native reference; two training sets name no licence

LANGUAGE	VERDICT	WHAT THE EVIDENCE SUPPORTS, IN ONE LINE	WHERE IT IS BEHIND, IN ONE LINE
Sanskrit	NO CLAIM	One small file for spoken Sanskrit and English; no comparison	IndicConformer Sanskrit; thin tests; chant and recitation not tested
Punjabi	WEAK CLAIM	Level with IndicConformer Punjabi and SraVaani-1.0 on mixed conversations	Both well ahead on read Punjabi and the native reference
Bodo	CLAIM	About 27% fewer errors than IndicConformer Bodo on 112 conversation turns	IndicConformer Bodo on the native reference and IndicVoices valid; one rival measured
Dogri	NO CLAIM	One small file for Dogri and English; no comparison	Conversation test of 24 turns; IndicConformer Dogri ahead on the one sizeable set
Kashmiri	WEAK CLAIM	Fewer errors than IndicConformer Kashmiri on IndicVoices valid only, a set that flatters it	The independent test is 32 turns with no English
Konkani	CLAIM	About 15% fewer errors than IndicConformer Konkani on 90 conversation turns	IndicConformer Konkani on the native reference and IndicVoices valid; Devanagari only
Maithili	NO CLAIM	One small file for Maithili and English; no comparison	IndicConformer Maithili ahead or level on every test
Manipuri	WEAK CLAIM	Fewer errors than IndicConformer Manipuri on 44 conversation turns, under our floor	IndicConformer Manipuri on the native reference and IndicVoices valid; Meetei Mayek only
Nepali	NO CLAIM	One small file for Nepali and English; no comparison with IndicConformer Nepali	IndicConformer Nepali well ahead on read Nepali; automatic language choice misfires
Santali	WEAK CLAIM	No clear difference from IndicConformer Santali on two tests	No lead over the one rival measured; thin English test; Ol Chiki only
Sindhi	NO CLAIM	One small file for Sindhi in Devanagari and English; no comparison	Devanagari only; conversation test of 27 turns

6. Language by language

All tables give word error rate, %, lower is better, on the same clips and with the same scorer within each table. "DiarBench, Latin" is the English-in-Latin reference and "DiarBench, native" is the native reference (section 4.3). "English in Latin" is the share of DiarBench's English words written in English letters. A dash means not measured or not reported in the facts file.

6.1 Shruti Hindi

Hindi in Devanagari, English in lowercase Latin. 29.5 MB, 55M parameters. Built on Cactus Whistle. Measured 18-19 October 2026. **Verdict: no label in its facts file.**

MODEL	FILE	MONSOON HINDI (LEADERBOARD SCORER, PUBLIC SPLIT)	FLEURS HINDI	LAHAJA (ACCENTED HINDI)	MUCS HINGLISH	SVARAH (INDIAN ENGLISH)	LIBRISPEECH	WORDS ON 100 SILENT CLIPS
Shruti Hindi	29.5 MB	20.46	19.92	26.28	35.21	22.47	5.95	0
Koyal-hi-120M	481 MB	11.64	9.56	18.70	57.51	Hindi only	Hindi only	0
IndicConformer Hindi	482 MB	15.31	11.31	13.18	64.48	Hindi only	Hindi only	88
Vosk small (hi + en-in)	139 MB	35.75	18.07	26.60	61.14	38.51	30.18	1
Oriserve Whisper-Hindi2Hinglish-Swift	296 MB	57.16 (Roman script only)	–	–	–	20.16	12.60	1,676
Whisper-small	971 MB	128.25	59.42	117.5	–	–	–	–

Our Monsoon number is on the public split and the leaderboard's own number may differ by a few points. We have made no official submission, so there is no rank.

Against Whisper large-v3, on the same clips (Mac, fp16, beam 5, language set; 3,087 MB; interval is Whisper minus Shruti Hindi):

SET (CLIPS)	SHRUTI HINDI	WHISPER LARGE-V3	DIFFERENCE, 95% INTERVAL
Monsoon Hindi (753)	20.46	33.29	+12.8, no interval on this scorer
FLEURS Hindi (418)	20.14	28.88	+8.7 (+6.0 to +12.5)
Lahaja (first 300)	22.87	31.19	+8.3 (+4.4 to +14.7)
MUCS Hinglish (first 600)	45.3	79.94	+34.6 (+23.5 to +48.7), partly script
Svarah, Indian English (first 300)	21.35	7.39	-14.0, Whisper far better
LibriSpeech (first 300)	5.76	2.25	-3.5, Whisper better
Words on 100 silent clips	0	538	

Plain beam search without Whisper's temperature fallback causes some repetition loops. With those clips removed, Shruti Hindi still leads (FLEURS 19.9 vs 26.4, Lahaja 22.9 vs 28.7). Shruti Hindi's training data includes FLEURS's training split, so Lahaja, which was kept out of training, is the cleanest Hindi comparison.

Hinglish (MUCS 2021, 3,136 clips): 35.22 vs whisper-base-hinglish 87.05 and whisper-tiny-hinglish-v2 87.73. Those two write Roman script only, so that mostly measures script. With both sides romanised: 34.32 vs 43.45 and 44.94. With the most forgiving romaniser the lead is narrow: 31.37 vs 32.66 and 34.22. Their sizes as measured are 290 MB and 151 MB (fp32). Their int8 builds (76 and 41 MB) were not measured, and we do not compare with the 28.4 and 32.0 on their card.

Conversations (DiarBench, 200 turns, 57 conversations, 1,553 English words; scored 7 November 2026 from saved transcripts). With English in Latin, Shruti Hindi scores 25.97 against SraVaani-1.0's 30.65 (908.8 MB) and IndicConformer Hindi int8's 39.10 (137.7 MB). Paired intervals: vs SraVaani-1.0 -7.15 to -2.42, by conversation -8.79 to -0.10, script-blind -2.97 to +0.84; vs IndicConformer Hindi -16.86 to -9.89, by conversation -21.23 to -5.59, script-blind -9.92 to -4.54. It writes 68.6% of the English words in English letters. So the lead over SraVaani-1.0 rests on writing English in English letters. On the native reference both are ahead: 40.06 vs 21.71 and 29.96.

The claim: the smallest open model we know of that does Hindi, English and Hinglish in one file, by file size. The lowest Hinglish word error rate of the open models we measured (MUCS 2021, same harness), still ahead when both sides are written in Roman script. About 100x smaller than Whisper large-v3 and more accurate on the Hindi test sets named above; Whisper is better on English.

Where it falls short: Koyal and IndicConformer Hindi are far better on pure Hindi. Whisper large-v3 is far better on English. Far-away or speaker-played audio can throw it: most Hindi and Hinglish clips played from a speaker across a desk were detected as English and came out wrong. Names it never heard in training are spelt by sound.

6.2 Shruti Marathi

Marathi in Devanagari, English in lowercase Latin. It does not do Hindi. 29.7 MB, 55,722,313 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 2 November 2026; rivals 30 October and 2 November. **Verdict: no label in its facts file.**

MODEL	FILE	FLEURS (READ)	KATHBATH VALID (READ)	DIARBENCH, LATIN	DIARBENCH, NATIVE	ENGLISH IN LATIN
Shruti Marathi	29.7 MB	29.9	24.5	38.2	40.9	40.5%
Whisper tiny	44 MB	273.1	278.1	373.9	389.7	–
Whisper base	82 MB	177.9	193.2	402.9	420.5	–
Dolphin-base	103.7 MB	71.6	74.0	97.0	97.8	–
33M MoE conformer (fp32)	136.7 MB	41.0	32.2	76.7	–	0%
IndicConformer Marathi int8	137.7 MB	22.4	14.1	41.5	32.3	0%

Intervals for Shruti Marathi, 95% bootstrap: FLEURS 27.9-32.2, Kathbath valid 21.8-27.6, DiarBench Latin 35.6-41.0. Larger models, same harness: Hima-ASR 0.6B (1,174 MB) FLEURS 18.2, Kathbath 15.0; IndicConformer-600M (2,560 MB) FLEURS 18.4, Kathbath 13.2; whisper-small-marathi (967 MB) FLEURS 55.1, Kathbath 54.0.

Other sets, Shruti Marathi vs IndicConformer Marathi: SPRING-INX R2 (code-mixed phone calls, in-domain for us) 42.9 vs 53.0; Vaani with English in Latin 32.2 vs 44.0, native 41.7 vs 37.6; IndicVoices valid native 26.6 vs 21.3. IndicVoices valid with English in Latin is 22.2 vs 27.6, from an older scoring of a reference type that later languages bar from claims, so we do not rest a claim on it. These sets share corpora with our training data. FLEURS and DiarBench are the held-out ones.

Marathi's DiarBench scores use its original references, which kept an extra word in 127 of the 260 Latin references. Scored on 7 November 2026 with the later references used for the other languages, Shruti Marathi is 35.3 (32.7-38.4) against IndicConformer Marathi's 41.1 with English in Latin, and 40.1 against 31.4 on the native reference. IndicConformer-600M scores 34.4 and 21.9. Hima-ASR 0.6B (1,174 MB, with a Gemma-licensed decoder) scores 21.8 and 32.1 in its mixed build and 34.1 and 21.4 in its native build. That is 200 turns, 42 conversations and 805 English words. Paired intervals with English in Latin, on those references: vs IndicConformer Marathi -8.20 to -3.38, by conversation -8.97 to -1.44, script-blind -6.87 to -1.76; vs IndicConformer-600M -1.25 to +3.34, so no lead; vs Hima-ASR's mixed build +11.16 to +15.88. The tables here keep the original scoring, as released.

The claim: the most accurate open speech-to-text model under 100 MB for Marathi that we measured, as of 2 November 2026 (FLEURS and Kathbath above). On conversations that mix

Marathi and English, fewer errors than IndicConformer Marathi when English words are written in English letters (DiarBench 38.2 vs 41.5). Within 1.33x of IndicConformer Marathi's error rate on FLEURS, at a fifth of its size.

Where it falls short: IndicConformer Marathi makes about a quarter fewer errors on FLEURS and about 40% fewer on Kathbath, and it is ahead on native references (DiarBench 32.3 vs 40.9). The 0.6B models are ahead by more. Hima-ASR's mixed build is well ahead even with English in Latin, 21.8 vs 35.3 on the later references, and it writes more English words in English letters, 65.2% vs 42.4%. Spelling variants count as errors, English loanwords sometimes come out in Devanagari, and the letters of acronyms can split.

6.3 Shruti Tamil

Tamil in Tamil script, English in lowercase Latin. 29.7 MB, 55,875,913 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 2 November 2026. **Verdict: no label in its facts file.**

MODEL	FILE	FLEURS (READ)	KATHBATH (READ)	DIARBENCH, LATIN	DIARBENCH, NATIVE	ENGLISH IN LATIN
Shruti Tamil	29.7 MB	38.5	33.2	52.5	62.2	51.5%
Whisper tiny	43.5 MB	226.7	239.6	244.8	–	–
Whisper base	81.8 MB	124.5	108.8	152.5	–	–
Thedal v11 (non-commercial licence)	102.6 MB	42.2	38.1	70.1	64.9	–
Dolphin-base int8	103.7 MB	71.9	65.0	92.4	–	–
IndicConformer Tamil int8	137.7 MB	27.8	25.0	58.7	51.4	0%
deepdml whisper-tiny-ta-mix-norm (fp32)	151.1 MB	47.9	45.6	90.3	–	–
deepdml whisper-base-ta-mix-norm (fp32)	290.4 MB	38.7	39.1	81.1	–	–
SraVaani-1.0	908.8 MB	24.5	21.5	54.8	44.0	0%

Every model other than Shruti Tamil writes 8% or less of DiarBench's English words in English letters.

Paired intervals on DiarBench with English in Latin (200 turns, 50 conversations, 987 English words; 7 November 2026, from saved transcripts): vs IndicConformer Tamil -8.94 to -3.55, by

conversation -8.84 to -2.61, script-blind -6.08 to -1.18; vs SraVaani-1.0 -4.90 to +0.45, by conversation -6.41 to +0.70, script-blind -3.46 to +1.70 (level).

The claim: the most accurate open Tamil speech-to-text model under 100 MB that we measured, as of 2 November 2026. On FLEURS it is level with deepdml whisper-base-ta (38.5 vs 38.7, paired interval -2.7 to +2.3), which is 290 MB as published. On conversations that mix in English, fewer errors than IndicConformer Tamil, when English words are written in English letters (DiarBench 52.5 vs 58.7, held out for both).

Where it falls short: IndicConformer Tamil makes about 28% fewer errors on FLEURS and 25% fewer on Kathbath. SraVaani-1.0 makes fewer still, and on the native reference too (44.0 vs 62.2). Language detection can pick English for Tamil speech with many English words: one live take of five came out garbled. Number words can be swapped.

6.4 Shruti Bengali

Bengali in Bengali script, English in lowercase Latin. 29.7 MB, 55,875,913 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 2 November 2026; int8 Vosk on 3 November. **Verdict: no label in its facts file.**

MODEL	FILE	FLEURS (READ)	KATHBATH (READ)	DIARBENCH, LATIN	DIARBENCH, NATIVE	ENGLISH IN LATIN
Shruti Bengali	29.7 MB	29.9	22.2	36.0	41.8	49.7%
Vosk Bengali int8 (third-party copy)	28.6 MB	20.9	22.4	43.3	39.1	0%
Whisper tiny	43.5 MB	361.6	403.7	540.9	543.2	15.2%
Whisper base	81.8 MB	176.7	214.5	339.9	341.6	11.7%
Vosk Bengali (Zipformer)	90.0 MB	20.8	21.65	42.8	38.5	0%
Dolphin-base int8	103.7 MB	41.1	45.9	93.0	93.2	2.7%
IndicConformer Bengali int8	137.7 MB	20.65	12.07	39.3	34.8	0%
Dolphin-small int8	249.7 MB	40.3	33.6	99.4	99.4	0.3%
Whisper small	264.5 MB	184.3	173.2	167.1	167.9	5.5%
SraVaani-1.0	908.8 MB	15.7	11.3	34.6	27.1	1.6%

Paired intervals on DiarBench with English in Latin: vs Vosk Bengali -9.3 to -4.3; vs IndicConformer Bengali -5.5 to -1.0; vs SraVaani-1.0 -0.9 to +3.8 (level). Stricter checks (200 turns, 64 conversations, 770 English words; 7 November 2026, from saved transcripts): vs IndicConformer Bengali by conversation -5.90 to -0.60, script-blind -4.15 to +0.37; vs SraVaani-1.0 by conversation -1.76 to +4.82, script-blind -0.57 to +4.09.

The claim: on real Bengali conversations that mix in English, with English words in English letters, fewer errors than Vosk Bengali (90 MB) and IndicConformer Bengali (137.7 MB). It writes about half of the English words in English letters, where those two write none. Level with Vosk Bengali on Kathbath (paired interval -1.6 to +2.7).

Where it falls short: a 28.6 MB int8 copy of Vosk Bengali, slightly smaller than this file, is clearly more accurate on read Bengali. Vosk Bengali makes about 30% fewer errors on FLEURS, and IndicConformer Bengali about 31% fewer on FLEURS and 46% fewer on Kathbath. SraVaani-1.0 is ahead on every set. Both named rivals are ahead on the native reference. The lead over IndicConformer Bengali does not hold when script is ignored, so it depends on writing English in

English letters. Whisper's rates are above 100% here, but Vosk Bengali is under 100 MB too, so this model does not lead its size class.

6.5 Shruti Telugu

Telugu in Telugu script, English in lowercase Latin. 29.7 MB, 55,875,913 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 2 November 2026. **Verdict: no label in its facts file.**

MODEL	FILE	FLEURS (READ)	KATHBATH (READ)	DIARBENCH, LATIN	DIARBENCH, NATIVE	ENGLISH IN LATIN
Shruti Telugu	29.7 MB	38.0	32.9	45.4	58.0	53.7%
Whisper tiny	43.5 MB	687.8	583.3	659.1	665.9	19.4%
Vosk small te	60.5 MB	77.1	80.1	93.8	92.0	0%
Whisper base	81.8 MB	220.1	363.3	371.3	375.5	11.9%
Dolphin-base int8	103.7 MB	73.4	66.2	86.9	90.3	14.2%
IndicConformer Telugu int8	137.7 MB	24.5	23.0	52.9	37.3	0%
vasista22 Whisper-tiny Telugu (fp32)	151.1 MB	34.2	34.4	59.2	47.5	0%
Dolphin-small int8	249.7 MB	60.9	52.6	98.0	98.7	2.6%
Whisper small	264.5 MB	169.7	154.1	201.1	205.1	14.1%
vasista22 Whisper-base Telugu	290.5 MB	28.9	29.5	55.3	41.3	0%
SraVaani-1.0	908.8 MB	22.0	20.3	47.8	33.0	4.6%
vasista22 Whisper-small Telugu	967.1 MB	24.0	26.0	55.7	42.7	0%

Paired intervals on DiarBench with English in Latin: vs IndicConformer Telugu -10.7 to -4.5; vs vasista22 Whisper-small -13.5 to -7.1; vs SraVaani-1.0 -5.8 to +0.5 (level); vs Vosk small te -51.0 to -45.4. Stricter checks (200 turns, 50 conversations, 1,016 English words; 7 November 2026, from saved transcripts): vs IndicConformer Telugu by conversation -12.30 to -3.93, script-blind

-5.72 to -0.13, which only just excludes 0; vs SraVaani-1.0 by conversation -6.89 to +2.04, script-blind -1.11 to +4.60.

The claim: on Telugu conversations that mix in English, with English words in English letters, fewer errors than IndicConformer Telugu, a model almost five times its size, and Whisper Telugu fine-tunes up to 32 times its size. Level with SraVaani-1.0 on those conversations. About half the errors of Vosk small te on every test we ran.

Where it falls short: IndicConformer Telugu makes about 36% fewer errors on FLEURS and 30% fewer on Kathbath. SraVaani-1.0 is ahead on every read set. The vasista22 Whisper-tiny fine-tune is ahead on FLEURS, and would be about 44 MB as a q8_0 copy. Most larger rivals are ahead on the native reference (SraVaani-1.0 33.0, IndicConformer Telugu 37.3, Shruti Telugu 58.0).

6.6 Shruti Gujarati

Gujarati in Gujarati script, English in lowercase Latin. 29.7 MB, 55,875,913 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 2 November 2026. **Verdict: no label in its facts file.**

MODEL	FILE	FLEURS (READ)	KATHBATH (READ)	DIARBENCH, LATIN	DIARBENCH, NATIVE	ENGLISH IN LATIN
Shruti Gujarati	29.7 MB	27.9	20.7	28.5	33.6	50.4%
Whisper tiny	43.5 MB	205.9	222.9	469.1	472.4	24.0%
Whisper base	81.8 MB	191.1	189.4	278.6	280.9	17.3%
Dolphin-base int8	103.7 MB	74.7	64.0	77.6	79.5	14.2%
Vosk small gu	108.1 MB	36.5	28.0	71.9	70.2	0%
IndicConformer Gujarati int8	137.7 MB	18.8	14.5	44.9	40.2	0%
Dolphin-small int8	249.7 MB	61.6	47.3	92.9	93.4	3.6%
Whisper small	264.5 MB	131.9	136.0	131.8	132.4	4.2%
vasista22 Whisper-base	290.5 MB	32.3	29.8	72.5	70.0	0%
SraVaani-1.0	908.8 MB	16.1	12.5	32.4	22.8	0%
vasista22 Whisper-small	967.1 MB	26.8	35.1	65.4	62.3	0%

Paired intervals on DiarBench with English in Latin: vs SraVaani-1.0 -5.66 to -2.14, by conversation -6.23 to -1.29, script-blind -4.68 to -1.21; vs IndicConformer Gujarati -19.03 to -13.91, by conversation -19.89 to -11.93, script-blind -16.82 to -11.75.

The claim: on Gujarati conversations that mix in English, with English words in English letters, fewer errors than every model we measured, including SraVaani-1.0. Also fewer errors than IndicConformer Gujarati against the native reference (33.6 vs 40.2). Fewer errors than Vosk small gu on every test we ran.

Where it falls short: IndicConformer Gujarati makes about 33% fewer errors on FLEURS and 30% fewer on Kathbath. SraVaani-1.0 is ahead on every read set and on the native reference (22.8 vs 33.6). vasista22 Whisper-small is level on FLEURS.

6.7 Shruti Odia

Odia in Odia script, English in lowercase Latin. 29.7 MB, 55,875,913 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 2 November 2026. **Verdict: CLAIM.**

MODEL	FILE	FLEURS (READ)	KATHBATH (READ)	DIARBENCH, LATIN	DIARBENCH, NATIVE	ENGLISH IN LATIN
Shruti Odia	29.7 MB	39.5	32.5	36.2	49.2	55.8%
Dolphin-base int8	103.7 MB	68.1	52.7	97.5	99.4	8.0%
IndicConformer Odia int8	137.7 MB	25.9	21.1	49.9	40.7	0.0%
Dolphin-small int8	249.7 MB	71.1	40.4	99.9	99.9	0.3%
SraVaani-1.0	908.8 MB	22.1	21.4	43.4	34.9	0.0%

Whisper's language list has no Odia. Paired intervals on DiarBench with English in Latin: vs SraVaani-1.0 -9.75 to -4.76, by conversation -10.72 to -1.20, script-blind -7.42 to -2.94; vs IndicConformer Odia -16.36 to -11.38, by conversation -16.29 to -8.55, script-blind -13.31 to -8.52.

The claim: on Odia conversations that mix in English, with English words in English letters, fewer errors than every model we measured. About 27% fewer than IndicConformer Odia and about 17% fewer than SraVaani-1.0, and both hold with conversations resampled and with script ignored. Fewer errors than Dolphin-base and Dolphin-small on every test we ran.

Where it falls short: read Odia is the weak side. IndicConformer Odia makes about 35% fewer errors on both read sets, and SraVaani-1.0 is ahead on both. Character errors on FLEURS are 18.38% vs 7.64%. Against the native reference, IndicConformer Odia (40.7) and SraVaani-1.0 (34.9) are ahead of 49.2. About 303 of 1,459 English words come out in Odia script.

6.8 Shruti Kannada

Kannada in Kannada script, English in lowercase Latin. 29.7 MB, 55,875,913 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 2 November 2026. **Verdict: CLAIM.**

MODEL	FILE	FLEURS (READ)	KATHBATH (READ)	DIARBENCH, LATIN	DIARBENCH, NATIVE	ENGLISH IN LATIN
Shruti Kannada	29.7 MB	33.1	29.4	56.6	69.5	53.1%
Whisper tiny	43.5 MB	305.5	371.7	574.5	584.5	23.0%
Whisper base	81.8 MB	179.5	178.5	248.8	252.4	8.2%
IndicConformer Kannada int8	137.7 MB	20.4	19.8	67.1	56.7	0.0%
vasista22 Whisper-tiny	151.1 MB	29.7	34.2	91.4	90.2	0.0%
Whisper small	264.5 MB	103.0	100.8	118.9	120.5	5.4%
vasista22 Whisper-base	290.5 MB	25.6	31.0	90.9	89.2	0.0%
SraVaani-1.0	908.8 MB	19.1	15.3	60.7	46.4	0.0%
vasista22 Whisper-small	967.1 MB	20.8	26.5	88.7	85.8	0.0%

Paired intervals on DiarBench with English in Latin: vs IndicConformer Kannada -13.21 to -7.82, by conversation -16.04 to -6.06, script-blind -9.00 to -4.40. Vs SraVaani-1.0 -7.17 to -1.25 by clip, but -11.19 to +1.64 by conversation and -2.87 to +2.34 script-blind, so we do not claim it: the whole gap is English letters.

The claim: on Kannada conversations that mix in English, with English words in English letters, fewer errors than IndicConformer Kannada, a model almost five times its size, and the three vasista22 Whisper Kannada fine-tunes, up to 33 times its size. About 16% fewer than IndicConformer Kannada.

Where it falls short: more than half of the words are wrong on real conversation: 56.6% with English in Latin and 69.5% against the native reference. A reader who tries it on real talk will see this. On FLEURS, IndicConformer Kannada, SraVaani-1.0 and all three vasista22 fine-tunes are ahead. Character errors on FLEURS are 12.69% vs 5.99%. About 248 of 1,091 English words come out in Kannada script.

6.9 Shruti Urdu

Urdu in Perso-Arabic script, right to left, English in lowercase Latin. 29.7 MB, 55,875,913 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 2 November 2026. **Verdict:**

CLAIM.

MODEL	FILE	FLEURS (READ)	KATHBATH (READ)	DIARBENCH, LATIN	DIARBENCH, NATIVE	ENGLISH IN LATIN
Shruti Urdu	29.7 MB	33.5	18.2	22.9	31.3	64.1%
Whisper tiny	43.5 MB	108.8	107.4	257.3	255.8	0.0%
Whisper base	81.8 MB	55.6	55.3	111.7	111.0	6.4%
Dolphin-base int8	103.7 MB	63.0	57.5	98.5	99.0	3.2%
IndicConformer Urdu int8	137.7 MB	24.7	13.4	29.1	19.5	0.0%
Dolphin-small int8	249.7 MB	75.3	51.3	98.5	99.0	3.2%
Whisper small	264.5 MB	41.7	45.7	50.8	50.0	10.8%

Paired intervals on DiarBench with English in Latin: vs IndicConformer Urdu -10.22 to -1.64, by conversation -9.99 to -1.51. The script-blind check does not work for this script.

The claim: on Urdu conversations that mix in English, counting English words only in English letters, about 21% fewer errors than IndicConformer Urdu, on all 74 usable turns from 25 conversations. Fewer errors than Dolphin base and small and Whisper tiny, base and small on every test we ran.

Where it falls short: the lead rests on English letters. Against the native reference IndicConformer Urdu is well ahead on the same turns (19.5 vs 31.3), and we cannot show a gain beyond English. On read Urdu it makes about 26% fewer errors on both sets. Character errors on DiarBench's native reference are 28.69% vs 7.40%. With 74 turns, the size of the lead is not pinned down. No set here is checked to be Pakistani speakers, so we can say only what was measured. A 43.9 MB Urdu Zipformer is gated and unmeasured, so this report makes no size comparison.

6.10 Shruti Malayalam

Malayalam in Malayalam script, English in lowercase Latin. 29.7 MB, 55,875,913 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 2 November 2026. **Verdict: WEAK CLAIM.**

MODEL	FILE	FLEURS (READ)	KATHBATH (READ)	DIARBENCH, LATIN	DIARBENCH, NATIVE	ENGLISH IN LATIN
Shruti Malayalam	29.7 MB	36.0	44.9	56.8	65.4	39.1%
Whisper tiny	43.5 MB	509.4	910.2	555.0	559.6	15.9%
Whisper base	81.8 MB	186.3	275.1	217.8	219.4	6.6%
IndicConformer Malayalam int8	137.7 MB	31.6	35.0	61.7	53.2	0.0%
Whisper small	264.5 MB	316.8	251.8	238.9	244.7	27.1%
SraVaani-1.0	908.8 MB	22.6	30.4	55.2	45.2	4.7%

Paired intervals on DiarBench with English in Latin: vs IndicConformer Malayalam -8.28 to -1.79 by clip, but -9.45 to +1.23 by conversation and -4.48 to +1.50 script-blind; vs SraVaani-1.0 -1.29 to +4.81 by clip, and +0.43 to +6.27 script-blind.

The claim: Malayalam and English in one 29.7 MB file that runs on your device. No comparison in its headline. Measured as written, the conversation lead over IndicConformer Malayalam passes by clip and fails the two stricter checks. It is level with SraVaani-1.0 as written, and SraVaani-1.0 is ahead with script ignored. On read Malayalam (FLEURS) it makes about 14% more errors than IndicConformer Malayalam, at about a fifth of the size. What would create an edge: a lead that holds with conversations resampled and script ignored.

Where it falls short: 56.8% of words are wrong on conversations with English in Latin, and 65.4% against the native reference. It has the lowest English-letter rate of the single launches: 39.1%, with about 336 of 1,038 English words in Malayalam script. Kathbath is its weakest read result, and word scores are harsh on a language that fuses words: 44.9% of words and 9.02% of characters on Kathbath.

6.11 Shruti Assamese

Assamese in Assamese script (Bengali-Assamese block), English in lowercase Latin. 29.7 MB, 55,875,913 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 2 November 2026; the int8 conversion on 6 November. Kathbath has no Assamese. Verdict: CLAIM.

MODEL	FILE	FLEURS (READ)	DIARBENCH, LATIN	DIARBENCH, NATIVE	ENGLISH IN LATIN
Shruti Assamese	29.7 MB	35.1	30.2	43.2	57.3%
Whisper tiny	43.5 MB	315.7	255.8	262.1	34.9%
Whisper base	81.8 MB	191.9	328.1	337.1	48.5%
IndicConformer Assamese int8 (community conversion)	198.8 MB	24.28	42.52	35.28	0.0%
SraVaani-1.0	908.8 MB	21.9	37.9	29.1	8.5%
IndicConformer-600M	2,560 MB	20.2	39.8	25.9	0.0%

Whisper small was not run on Assamese. Paired intervals on DiarBench with English in Latin: vs IndicConformer-600M -16.37 to -3.74, by conversation -15.81 to -3.30, script-blind -15.20 to -3.18; vs SraVaani-1.0 -13.26 to -3.11, by conversation -12.20 to -3.59, script-blind -12.53 to -2.76; vs the int8 conversion -18.66 to -6.66, by conversation -18.16 to -6.97, script-blind -15.80 to -4.54.

The claim: on Assamese conversations that mix in English, with English words in English letters, fewer errors than IndicConformer-600M and SraVaani-1.0, on the 84 usable turns from 19 conversations. Also fewer errors than a 198.8 MB int8 conversion of the per-language IndicConformer Assamese. That conversion is a community one, not AI4Bharat's own checkpoint, and about a third of its transcripts hold doubled vowel signs. We have not run the original.

Where it falls short: two training sets name no licence (section 9). Read Assamese is a wide miss: FLEURS 35.1 vs 20.2 and 21.9. Against the native reference, 43.2 vs 25.9 and 29.1, and character errors are 26.69% vs 9.27% for SraVaani-1.0. On the engine's own language choice, 3 of 200 FLEURS clips and 3 of 84 DiarBench turns came out wholly in Latin letters. There is only one read test.

6.12 Shruti Sanskrit

Spoken Sanskrit in Devanagari, English in lowercase Latin. 29.7 MB, 55,722,313 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 3 November 2026; Su-śrotā on 6 November. **Verdict: NO CLAIM.**

MODEL	FILE	INDICVOICES VALID	DIARBENCH, LATIN	DIARBENCH, NATIVE	ENGLISH IN LATIN
Shruti Sanskrit	29.7 MB	30.9	32.3	38.1	61.8% of 68 words
Whisper tiny	43.5 MB	617.4	522.9	528.1	48.5%
Whisper base	81.8 MB	440.6	367.7	374.7	69.1%
IndicConformer Sanskrit int8	137.7 MB	24.4	36.6	33.4	0.0%
Su-śrotā int8 ONNX	187.5 MB	38.7	43.8	41.8	0.0%
Whisper small	264.5 MB	443.3	367.4	369.4	20.6%

The claim: none against a serious rival. Plain line: spoken Sanskrit and English in one 29.7 MB file that runs on your device. IndicConformer Sanskrit is ahead on IndicVoices valid (+4.29 to +9.01). On DiarBench's 55 turns the Latin-reference interval is 14 points wide (-11.44 to +2.77), which is no evidence either way. Su-śrotā writes words joined more often, and each join counts as errors; on characters it is level or ahead (DiarBench native 12.13% vs our 15.90%), so we make no claim against it. Whisper is mostly writing nonsense or Latin on these tests, so those numbers say little.

Where it falls short: thin tests: no FLEURS, no Kathbath, one source shared with our training data, and 55 turns of conversation with only 68 English words. IndicConformer Sanskrit's character errors are about half of ours (6.10% vs 11.75%). Chant, verse and recitation were not tested. What would create an edge: an independent read set, more speech, then a lead over IndicConformer Sanskrit that holds on characters as well as words.

6.13 Shruti Punjabi

Punjabi in Gurmukhi, English in lowercase Latin. 29.7 MB, 55,875,913 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 2 November 2026. **Verdict: WEAK CLAIM.**

MODEL	FILE	FLEURS (READ)	KATHBATH (READ)	DIARBENCH, LATIN	DIARBENCH, NATIVE	ENGLISH IN LATIN
Shruti Punjabi	29.7 MB	29.3	18.6	32.6	38.3	58.7%
Whisper tiny	43.5 MB	311.5	337.8	468.7	467.7	13.4%
Whisper base	81.8 MB	201.6	207.9	275.9	275.3	6.1%
Dolphin-base int8	103.7 MB	83.6	71.6	98.5	99.4	7.6%
IndicConformer Punjabi int8	137.7 MB	16.9	13.7	34.6	28.4	0.0%
Dolphin-small int8	249.7 MB	62.2	51.0	99.2	99.2	0.2%
Whisper small	264.5 MB	129.4	123.4	145.1	144.7	0.6%
SraVaani-1.0	908.8 MB	15.1	12.7	31.9	23.9	0.0%

Paired intervals on DiarBench with English in Latin: vs IndicConformer Punjabi -4.44 to +0.15, by conversation -6.37 to +0.68, script-blind -2.69 to +1.53; vs SraVaani-1.0 -1.34 to +2.78, by conversation -2.30 to +3.06, script-blind -0.70 to +3.33. All include 0.

The claim: on Punjabi conversations, level with IndicConformer Punjabi and SraVaani-1.0, and it writes 58.7% of English words in English letters, where those two write none. Fewer errors than Dolphin base and small and Whisper tiny, base and small on every test we ran.

Where it falls short: it wins nothing against a serious rival. On FLEURS, IndicConformer Punjabi (16.9) and SraVaani-1.0 (15.1) are far ahead of 29.3. Against the native reference, 38.3 vs 28.4 and 23.9. This set has less English than most (659 English words), so English letters help less. Local names glossed in Latin count as errors.

6.14 Shruti Bodo

Bodo in Devanagari, English in lowercase Latin. 29.7 MB, 55,722,313 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 3 November 2026. **Verdict: CLAIM.**

MODEL	FILE	INDICVOICES VALID	DIARBENCH, LATIN	DIARBENCH, NATIVE	ENGLISH IN LATIN
Shruti Bodo	29.7 MB	34.5	39.9	51.5	50.4%
IndicConformer Bodo int8	137.7 MB	28.6	55.0	46.9	0.0%

Paired interval on DiarBench with English in Latin: -19.42 to -10.63, by conversation -19.89 to -10.54, script-blind -17.42 to -8.43.

The claim: on Bodo conversations, about 27% fewer errors than IndicConformer Bodo, a model almost five times its size, on all 112 usable turns from 25 conversations (97 with English words), with English in English letters. IndicConformer Bodo is the one rival we measured.

Where it falls short: IndicConformer Bodo is ahead on the native reference (46.9 vs 51.5, +1.77 to +7.65) and on IndicVoices valid (28.6 vs 34.5). Character errors on DiarBench are 28.99% vs 14.58%. There is no independent read set. Whisper, SraVaani-1.0, Dolphin and Vosk do not list Bodo, and a 967 MB Whisper Bodo model is unmeasured.

6.15 Shruti Dogri

Dogri in Devanagari, English in lowercase Latin. 29.7 MB, 55,722,313 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 3 November 2026. Verdict: NO CLAIM.

MODEL	FILE	INDICVOICES VALID	DIARBENCH, LATIN	DIARBENCH, NATIVE
Shruti Dogri	29.7 MB	43.5	34.5	37.0
IndicConformer Dogri int8	137.7 MB	38.5	39.7	37.4

The claim: none, and no comparison. Plain line: Dogri and English in one 29.7 MB file that runs on your device. It keeps English words in English letters, which we show by example and do not count, because DiarBench Dogri holds only 40 English words.

Where it falls short: the test cannot support a claim: 24 turns from 17 conversations, against floors of 50 turns and 200 English words. IndicConformer Dogri is ahead on IndicVoices valid (+2.58 to +7.27), the one sizeable set. It is the one rival measured; Whisper does not list Dogri. What would create an edge: a bigger independent test.

6.16 Shruti Kashmiri

Kashmiri in Perso-Arabic script, right to left, English in lowercase Latin. It writes no Devanagari Kashmiri. 29.7 MB, 55,875,913 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 3 November 2026. Verdict: WEAK CLAIM.

MODEL	FILE	INDICVOICES VALID	DIARBENCH (HOLDS NO ENGLISH)
Shruti Kashmiri	29.7 MB	43.0	56.4
IndicConformer Kashmiri int8	137.7 MB	48.7	58.2

The claim: weak only. On held-out IndicVoices Kashmiri, fewer errors than IndicConformer Kashmiri (-8.39 to -3.17 by clip). That set's speakers are kept out of our training but it is **the same source as our training data**, so it flatters both models. On that set it writes 75.6% of glossed English words in Latin letters (303 words).

Where it falls short: its one lead is on the set that shares a source with its training data. The independent check is 32 turns with no English words: 56.4 vs 58.2, interval -7.93 to +4.45, too small to agree or disagree. So no result about mixed conversation exists for Kashmiri. Character errors favour IndicConformer Kashmiri: IndicVoices valid 23.16% vs 15.71%, DiarBench 34.9% vs 25.01%. Newer models, `Aazu34/kashmiri-asr-ctc` (186.6 MB) among them, are unmeasured.

6.17 Shruti Konkani

Konkani in Devanagari, English in lowercase Latin. It does not write Romi or Kannada-script Konkani. 29.7 MB, 55,722,313 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 3 November 2026. **Verdict: CLAIM.**

MODEL	FILE	INDICVOICES VALID	DIARBENCH, LATIN	DIARBENCH, NATIVE	ENGLISH IN LATIN
Shruti Konkani	29.7 MB	41.4	40.0	47.3	49.6%
IndicConformer Konkani int8	137.7 MB	33.4	47.0	37.3	0.0%

Paired interval on DiarBench with English in Latin: -10.64 to -3.57, by conversation -10.86 to -3.50, script-blind -6.96 to -0.90.

The claim: on Konkani conversations, about 15% fewer errors than IndicConformer Konkani, a model almost five times its size, on all 90 usable turns from 23 conversations (71 with English words), with English in English letters. IndicConformer Konkani is the one rival we measured.

Where it falls short: Devanagari only; Goan Konkani is widely written in Romi too, and a reader may expect that. IndicConformer Konkani is ahead on the native reference (37.3 vs 47.3, +6.41 to +13.71) and on IndicVoices valid (33.4 vs 41.4). Character errors on DiarBench are 25.38% vs 11.58%. It has the least training speech of the 16 newer languages, and "so" came out in Devanagari 18 times.

6.18 Shruti Maithili

Maithili in Devanagari, English in lowercase Latin. 29.7 MB, 55,722,313 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 3 November 2026. **Verdict: NO CLAIM.**

MODEL	FILE	INDICVOICES VALID	DIARBENCH, LATIN	DIARBENCH, NATIVE	ENGLISH IN LATIN
Shruti Maithili	29.7 MB	42.5	45.7	51.2	46.8% of 109 words, 54 turns
IndicConformer Maithili int8	137.7 MB	39.3	46.9	40.2	0.0%

The claim: none. Plain line: Maithili and English in one 29.7 MB file that runs on your device.

Where it falls short: IndicConformer Maithili is ahead or level on every test. IndicVoices valid +1.01 to +5.73; DiarBench native +6.67 to +14.77. On DiarBench with English in Latin the interval is 10 points wide (-6.29 to +3.88), on 54 turns, which shows nothing either way. Character errors on DiarBench's native reference are 31.26% vs 15.02%. `rockerritesh/whisper-tiny-maithili` (151.1 MB, about 44 MB as q8_0) is unmeasured.

6.19 Shruti Manipuri

Manipuri in Meetei Mayek, English in lowercase Latin. It does not write Manipuri in Bengali script. 29.7 MB, 55,875,913 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 3 November 2026. Verdict: WEAK CLAIM.

MODEL	FILE	INDICVOICES VALID	DIARBENCH, LATIN	DIARBENCH, NATIVE	ENGLISH IN LATIN
Shruti Manipuri	29.7 MB	36.0	48.2	66.0	60.9% of 243 words
IndicConformer Manipuri int8	137.7 MB	31.2	65.4	53.5	0.0%

Paired interval on DiarBench with English in Latin: -26.78 to -7.98, by conversation -31.04 to -8.05. The script-blind check does not work for this script.

The claim: weak only. On 44 Manipuri conversation turns from 20 conversations (39 with English words), with English in English letters, fewer errors than IndicConformer Manipuri, a model almost five times its size. 44 turns is under our floor of 50, so the number travels with the claim.

Where it falls short: the interval is wide and the set is small. Against the native reference IndicConformer Manipuri is far ahead (53.5 vs 66.0), and character errors are 40.29% vs 17.48%. It is also ahead on IndicVoices valid (31.2 vs 36.0). Of 200 IndicVoices clips, 4 came out wholly in Latin letters, and 2 of 44 DiarBench turns did too. What would create an edge: 50 or more turns with the lead holding by conversation.

6.20 Shruti Nepali

Nepali in Devanagari, English in lowercase Latin. 29.7 MB, 55,722,313 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 3 November 2026. Verdict: NO CLAIM.

MODEL	FILE	FLEURS (READ)	INDICVOICES VALID	DIARBENCH, LATIN	DIARBENCH, NATIVE	ENGLISH IN LATIN
Shruti Nepali	29.7 MB	46.0	27.1	34.0	40.5	55.9% of 118 words, 45 turns
Whisper tiny	43.5 MB	270.5	488.6	308.4	310.5	17.8%
Whisper base	81.8 MB	200.4	371.4	287.2	288.3	7.6%
IndicConformer Nepali int8	137.7 MB	29.9	18.4	37.4	28.6	0.0%
Whisper small	264.5 MB	126.7	226.5	238.0	237.6	0.0%

The claim: none against IndicConformer Nepali. Plain line: Nepali and English in one 29.7 MB file that runs on your device. Fewer errors than Whisper tiny, base and small on every test we ran, though Whisper's rates are above 100%.

Where it falls short: the independent read test is poor: FLEURS 46.0 vs IndicConformer Nepali's 29.9 (+13.60 to +19.38). IndicVoices valid (+6.72 to +11.13) and DiarBench native (+8.07 to +16.62) also favour it. **Its automatic language choice misfires:** see section 8.2. DiarBench's 45 turns are too few to say anything against IndicConformer Nepali. `ailabpvtltd/indicconformer-nepali-english-codemix-asr` (513.5 MB, CC BY-NC 4.0) is built for this mix and reports WER 19.9 on its own set; we have not measured it. So "writes no English" holds only for stock IndicConformer Nepali.

6.21 Shruti Santali

Santali in Ol Chiki, English in lowercase Latin. It writes no other Santali script. 29.7 MB, 55,875,913 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 3 November 2026. **Verdict: WEAK CLAIM.**

MODEL	FILE	INDICVOICES VALID	DIARBENCH, LATIN	DIARBENCH, NATIVE
Shruti Santali	29.7 MB	40.3	33.0	33.6
IndicConformer Santali int8	137.7 MB	42.6	35.4	34.2

Paired intervals: DiarBench native -3.16 to +2.03, by conversation -2.67 to +1.61; DiarBench Latin -4.94 to +0.25; IndicVoices valid native -5.29 to +0.52. All include 0.

The claim: weak only. On our two tests we found no clear difference from IndicConformer Santali, a model almost five times its size. The DiarBench intervals are about 5 points wide on 99

turns, so "no clear difference" is a real statement here. The word "fewer" does not apply.

Where it falls short: no lead over the one rival measured; the size gap is the entire story. Character errors favour IndicConformer Santali: DiarBench 15.88% vs 11.7%. The English test is thin: 34 of 99 turns and 63 English words, and about 30 of those came out in Ol Chiki. What would create an edge: a real lead, or an English test above 200 words.

6.22 Shruti Sindhi

Sindhi in Devanagari, not Perso-Arabic, English in lowercase Latin. 29.7 MB, 55,722,313 parameters. Built on Cactus Whistle via Shruti Hindi. Measured 3 November 2026. **Verdict: NO CLAIM.**

MODEL	FILE	INDICVOICES VALID	DIARBENCH, LATIN	DIARBENCH, NATIVE
Shruti Sindhi	29.7 MB	33.0	34.2	38.0
IndicConformer Sindhi int8	137.7 MB	33.1	38.7	35.8
Whisper tiny (not a comparison, see below)	43.5 MB	227.3	139.6	139.9
Whisper base (not a comparison, see below)	81.8 MB	311.6	243.4	243.4
Whisper small (not a comparison, see below)	264.5 MB	387.3	175.8	177.1

The claim: none. Plain line: Sindhi, written in Devanagari, and English in one 29.7 MB file that runs on your device. The Whisper rows do not show accuracy: the references are in Devanagari and Whisper hardly writes it (base writes 100% Arabic script, tiny no Devanagari, small about 9-10%).

Where it falls short: Devanagari only. Sindhi readers who use Perso-Arabic cannot use the output, and for the same reason FLEURS Sindhi was left out, so no independent read test exists. DiarBench is 27 turns from 17 conversations with 51 English words, too small. The one level result, IndicVoices valid (-2.45 to +2.39), is on the set that shares a source with our training data. A Perso-Arabic build would be a different model.

---

7. English, silence, speed and memory

Phone is a Samsung Galaxy F62 (a five-year-old phone with 6 GB of RAM), offline, in the native engine. Shruti Hindi was measured on one 60-second recording, the others on 20 clips each, so its row is not directly comparable. Some runs were back-to-back sessions, and heat may have slowed them (Bengali's and Telugu's facts files say so).

LANGUAGE	LIBRISPEECH, ENGLISH SET	WORDS ON 100 SILENT CLIPS	TIMES FASTER THAN REAL TIME ON THE PHONE	PEAK MEMORY	AUDIO IN THAT RUN
Hindi	5.95	0	6.3x	80 MB	60 s recording
Marathi	6.23	0	12.1x	64 MB	212 s
Tamil	6.37	0	10.8x	64 MB	249 s
Bengali	6.30	0	12.6x	64 MB	20 clips
Telugu	6.22	0	9.8x	64 MB	20 clips
Gujarati	6.17	0	10.3x	64 MB	20 clips
Odia	6.25	0	10.3x	64 MB	226.5 s
Kannada	6.16	0	11.2x	64 MB	234.9 s
Urdu	6.07	0	11.1x	64 MB	20 clips
Malayalam	6.24	0	10.1x	64 MB	255.7 s
Assamese	6.22	0	10.7x	64 MB	212.0 s
Sanskrit	5.93	0	12.0x	64 MB	202.6 s
Punjabi	6.24	0	10.7x	64 MB	251.2 s
Bodo	5.99	0	10.9x	64 MB	148.5 s
Dogri	6.1	0	9.5x	64 MB	139.1 s
Kashmiri	6.08	0	8.6x	64 MB	120.2 s
Konkani	6.0	0	11.7x	64 MB	187.5 s
Maithili	6.04	0	9.1x	64 MB	112.0 s
Manipuri	5.97	0	11.2x	64 MB	135.9 s
Nepali	6.01	0	9.7x	64 MB	113.5 s
Santali	5.98	0	12.6x	64 MB	202.2 s
Sindhi	6.05	0	10.8x	64 MB	212.5 s

The silent-clip test covers silence. It says nothing about noise or music.

In a browser. On the same phone, in Chrome with the WebAssembly build, on a page served from a local build, with clips played into the page and no real microphone. "On the device" is the page's own timing, not the wait for final text.

LANGUAGE	LINES	SPEECH	ON THE DEVICE
Bengali	5	8.4-9.6 s	4.0-4.5 s
Telugu	5	6.8-9.0 s	2.9-4.5 s
Gujarati	5	7.2-8.3 s	2.9-3.6 s
Odia	4	7.6-10.0 s	3.2-4.1 s
Kannada	5	7.4-8.6 s	2.9-3.7 s
Urdu	5	6.8-8.6 s	2.9-3.4 s
Malayalam	9	6.5-7.5 s	1.1-1.6 s
Assamese	7	6.9-8.6 s	1.0-1.5 s
Sanskrit	5	12-14 s	1.9-2.3 s
Punjabi	9	6.7-8.3 s	1.1-1.3 s
Bodo (test clips, not real speech)	4	7.8-9.1 s	1.1-1.6 s
Dogri	4	8.8-9.2 s	1.5-1.6 s
Kashmiri (test clips, not real speech)	4	6.5-8.3 s	1.4-3.2 s
Konkani	4	8.3-8.9 s	3.1-3.9 s
Maithili	5	7.9-9.4 s	2.6-4.0 s
Manipuri (test clips, not real speech)	4	7.3-8.0 s	2.0-2.9 s
Nepali	5	7.8-10.0 s	3.1-4.8 s
Santali (test clips, not real speech)	4	6.4-9.4 s	2.4-3.9 s
Sindhi	4	7.1-9.3 s	2.7-3.8 s

These are small runs on one phone. Bengali to Urdu ran on 2 November, on an earlier page; the others on 6 November. Shruti Hindi: on the live page in the phone's Chrome, the final pass took about a third of the speech time (for example, 28 s took 8.2 s). Shruti Marathi: audio files of 3.0-5.0 s gave final text in 1.4-2.8 s. Shruti Tamil: audio files of 4.6-7.0 s took 3.0-3.8 s.

---

8. Family limits

8.1 Indian English on automatic language choice

The engine can choose the language itself or take it as a setting. On Shruti Hindi, automatic choice matches the set language on Hindi, Hinglish, accented Hindi and US English (Monsoon 20.47 vs 20.46, Lahaja 22.89 vs 22.87, LibriSpeech 6.04 either way). **Indian-accented English is**

its weak spot: on a Svarah subset, 21.13 with English set, 25.75 on automatic, 27.52 with Hindi set. Automatic choice often treats that speech as Hindi. For speech that is entirely English, set `en`. We measured this comparison on Shruti Hindi and have not repeated it in the other languages. For English alone, Whisper large-v3 is far more accurate (section 6.1).

8.2 Nepali's language setting

On Nepali the automatic choice misfires more than on any other language in the family: on FLEURS, 20 of 200 clips came out wholly in English letters. With Nepali set, none of the 200 did, and FLEURS moves from 46.0 to 42.2. That helps and closes part of the gap with IndicConformer Nepali (29.9), not all of it. The numbers in section 6.20 use the automatic choice. The try box on our site sets the language, and our demo sets it for Nepali. Library and command-line users get the automatic choice unless they set it.

8.3 Setting the language costs plain English a little

The site's try box sends the language you picked. That keeps speech in the right script, and it costs some accuracy on plain English. WER on 12 plain-English clips, automatic then set: Hindi 1.45 then 11.6, Odia 2.9 then 6.5, Kannada 0 then 3.6, Gujarati 1.45 then 5.8. Not measured on the other languages.

8.4 The numbers are high

On real conversation many of these models get a large share of words wrong. Malayalam (56.8), Kannada (56.6) and Tamil (52.5) are above 50% even on their favourable reference. Word error also overstates the damage on languages that fuse words: Malayalam on Kathbath is 44.9% of words and 9.02% of characters, and Sanskrit on IndicVoices valid is 30.9% and 11.75%. Spelling variants and joined words count as errors. If accuracy matters more than size, use a larger model.

8.5 Scripts

Each model writes one script for its language. Sindhi is Devanagari only, Konkani is Devanagari only, Manipuri is Meetei Mayek only, Santali is Ol Chiki only, and Kashmiri is Perso-Arabic only. English words do not always stay in English: among languages with 200 or more English words on DiarBench, the English-letter share runs from 39.1% (Malayalam) to 64.1% (Urdu). Many of the others come out in the native script.

8.6 What we have not tested

- Noise, music, far-away and speaker-played audio on these languages. Shruti Hindi's own phone test found that speaker-played audio confuses language detection.
- Loud noise on these languages. On the everyday-speech board, Shruti models sometimes repeat a phrase over and over in loud noise (9 of 3,400 clips at 5 dB, measured on Hindi-family runs).
- A live microphone, on most languages; the browser runs above played clips into the page.

- Linux, Windows and iOS, for the 21 newer models.
- **Reading by speakers.** The numbers and transcripts of the 16 newer languages have not been read by a speaker of those languages. The facts files of the six earlier languages do not report such a reading either.

There is no punctuation, and numbers come out as words. Rare names are spelt by sound. These models should not be used in medical, legal or similar high-stakes settings without a person checking the output.

8.7 Rivals we have not measured

Claims mean "the models we measured", and these are the unmeasured models most likely to matter. A card's own number is not comparable until measured.

LANGUAGE	NOT MEASURED
Bengali	`ehzawad/whistle-bn-BD` (30.5 MB, gated, licence "other")
Telugu	`srinikvs/maata-telugu-small` (264.5 MB, gated)
Kannada	Koyal-kn-120m (481 MB); three small Whisper-tiny Kannada copies (75.5-97.8 MB)
Urdu	`mahwizzzz/Urdu-Streaming-Zipformer-20M` (43.9 MB, gated)
Malayalam	`serendipitylabs/dulu-malayalam-asr`; Koyal-ml-120m (481 MB)
Assamese	AI4Bharat's original IndicConformer Assamese checkpoint
Sanskrit	`Harveenchadha/vakyansh-wav2vec2-sanskrit-sam-60` (377.8 MB)
Punjabi	`kingabzpro/wav2vec2-large-xlsr-53-punjabi` (377.8 MB)
Bodo	`COCO1033/bodo-whisper-asr-v1` (967 MB)
Kashmiri	`Aazu34/kashmiri-asr-ctc` (186.6 MB); the Moonshine adapters
Konkani	`Reubencf/konkani-whisper-stt` (adapter, Romi script)
Maithili	`rockerritesh/whisper-tiny-maithili` (151.1 MB, about 44 MB as q8_0)
Nepali	`ailabpvtltd/indicconformer-nepali-english-codemix-asr` (513.5 MB, CC BY-NC 4.0)

---

9. Licences and data

Model licence. Each model is Apache-2.0, keeping Cactus Compute's `NOTICE`. The model licence does not set or change the terms of any dataset. It covers the model, not its training data.

Engine. The Cactus needle engine is Apache-2.0 and is not part of our model repositories.

Training data by language. Every language other than Hindi also reuses English from Shruti Hindi's training data: LibriSpeech (CC BY 4.0), Mozilla Common Voice 17 English (CC0) and AI4Bharat NPTEL (CC BY 4.0).

LANGUAGE	SOURCES FOR THAT LANGUAGE
Hindi	AI4Bharat IndicVoices, ARTPARK-IISc Vaani, AI4Bharat Shrutilipi, AI4Bharat Kathbath, AI4Bharat Rural_Women_ASR_v2, Google FLEURS Hindi train split, MUCS 2021 Hinglish, plus AI4Bharat NPTEL, Common Voice 17 English (India and South Asia accents) and LibriSpeech
Marathi, Tamil, Bengali, Gujarati, Malayalam, Punjabi	IndicVoices, Kathbath, SPRING-INX R1 and R2, Vaani (transcribed part)
Odia, Kannada	IndicVoices, Kathbath, SPRING-INX R1, Vaani (transcribed part)
Telugu, Urdu	IndicVoices, Kathbath, Vaani (transcribed part)
Assamese	IndicVoices, SPRING-INX R1 and R2, Vaani (transcribed part)
Kashmiri, Konkani, Maithili, Manipuri, Nepali, Santali, Sindhi	IndicVoices, Vaani (transcribed part)
Sanskrit, Bodo, Dogri	IndicVoices

What the sources' licences are, and how far we checked them:

- **AI4Bharat IndicVoices:** CC BY 4.0, in the Hugging Face tag and the card text.
- **AI4Bharat Kathbath:** Hugging Face tag CC BY 4.0. Its card text also says the packaging is CC0 and that they do not own the raw text.
- **ARTPARK-IISc Vaani:** Hugging Face tag CC BY 4.0, with no licence text in the card body.
- **SPRING-INX (SPRING Lab, IIT Madras), other than Assamese:** CC BY 4.0 on the official AIKosh listing. The Hugging Face copies we trained on have no licence of their own. We have not verified that the Hugging Face copies hold the same data as the AIKosh files, or that SPRINGLab on Hugging Face is the lab itself.
- **SPRING-INX Assamese R1 and R2: these name no licence on their pages,** and we could not find an AIKosh listing for them. The SPRING-INX paper says "we are open sourcing SPRING-INX data" and names Assamese among its languages; that is not a named licence. We kept these two sets in training and do not claim a licence for them. Apache-2.0 on Shruti Assamese does not set or change their terms.
- **MUCS 2021 Hinglish,** via `dianavdavidson/MUCS-Hinglish`: CC BY 4.0 on the card; the original corpus is CC BY-SA 4.0.
- **Google FLEURS, Rural_Women_ASR_v2, Shrutilipi:** CC BY 4.0.

This is a statement of what we read on each page, not legal advice. Please credit the datasets if you build on these models.

Rivals' licences vary. IndicConformer's per-language models are MIT, through OpenVoiceOS's ONNX exports. SraVaani-1.0 is MIT and gated. Whisper is MIT, Dolphin Apache-2.0 and the vasista22 fine-tunes Apache-2.0. Thedal v11 is non-commercial.

---

10. ONNX and developer integration

There are two routes, and neither needs a key, an account or a server: the weights download with no sign-in, and transcription runs on the device. The ONNX packs fit the runtimes many developers already use. The Cactus needle engine runs the released `.cact` files. **Every score in section 6 is the engine's, not a pack's.**

10.1 The ONNX packs

All 22 languages also come as ONNX packs for ONNX Runtime, the browser and sherpa-onnx, under Apache-2.0, at udgamlabs/Shruti-ONNX. The 32 MB build holds the same 4-bit weights as the `.cact` file, in standard ONNX operators. The 53 MB downloads are the low-memory build and the sherpa-onnx phone pack.

YOU ARE BUILDING	USE	DOWNLOAD PER LANGUAGE	SET THE LANGUAGE?	TESTED ON
A Python app or a server	`pip install shruti-onnx`	32 MB, or 53 MB for low memory	Yes	Mac, Linux x86
A web page	`npm i shruti-onnx onnxruntime-web`	32 MB	Yes	Mac, Chromium-based browser, Hindi and Urdu
A Node app	`shruti-onnx`	32 MB	Yes	Mac, Linux x86
An Android app	sherpa-onnx pack	53 MB	No, always detected	Galaxy F62, all 22 languages, 20 to 30 clips each
A desktop app in C or C++	sherpa-onnx pack	53 MB	No, always detected	Mac
An iOS app	sherpa-onnx pack	53 MB	No, always detected	Not run on an iPhone

```
from shruti_onnx import Shruti

asr = Shruti.load("urdu")
print(asr.transcribe("clip.wav")["text"])
```

```
import * as ort from "onnxruntime-web";
import { createShruti } from "shruti-onnx";

const asr = await createShruti({ ort, language: "urdu" });
const { text } = await asr.transcribe(samples16k);   // Float32Array, 16 kHz
mono
```

10.2 Test results, language by language

We compared each pack with the released model in the needle engine, on that language's own test sets, under six rules written before any run:

1. All three builds export and load.
2. With the language set, each build's accuracy matches the engine on every set, English included, and it writes no words on silence.
3. The same text across runtimes: the phone pack matches Python, and Node matches Python, on every clip tested, in the right script.
4. No characters lost, reordered or replaced.
5. With the language detected, the phone pack's accuracy matches the engine on every set.
6. The phone pack writes no words on silence.

"Matches" means the paired 95% interval of the difference includes 0. MIXED means accuracy held on every main set and a named part did not. FAIL means that on at least one main set, a build missed on every reference scored there.

No language passed all six rules. 13 are MIXED and 9 FAIL, Hindi among them. No rule was changed after the runs. We publish all 22 with these results so you can judge your own language. If you want the behaviour section 6 measures, use the engine route (section 10.4). These tests compare the packs with the engine, not with the truth for each user, and they ran on 6 and 7 November 2026.

LANGUAGE	VERDICT	SETS (CLIPS)	ACCURACY MISSES AGAINST THE ENGINE	LARGEST MISS, POINTS (95% INTERVAL)	OTHER RULES	LINUX X86
Hindi	FAIL	DiarBench 200, Kathbath 200, FLEURS 200	32 MB: 2 worse; 53 MB: 2 worse; phone pack: 1 worse	+1.25 (+0.14 to +3.21), 53 MB, DiarBench, native	Pack and Python differ on 2 of 1,679 clips	FAIL
Marathi	MIXED	DiarBench 200, Kathbath 200, FLEURS 200	32 MB: 1 worse; 53 MB: 1 worse	+1.62 (+0.25 to +3.32), 32 MB, DiarBench, English in Latin	Pack and Python differ on 2 of 1,679 clips; English letter share differs by over 1 point	MIXED
Tamil	FAIL	DiarBench 200, Kathbath 200, FLEURS 200	53 MB: 1 worse	+0.71 (+0.15 to +1.29), 53 MB, FLEURS, native	Pack and Python differ on 2 of 1,679 clips	FAIL
Bengali	MIXED	DiarBench 200, Kathbath 200, FLEURS 200	phone pack: 1 worse	+0.90 (+0.21 to +1.66), phone pack, Svarah, language detected	Pack and Python differ on 3 of 1,679 clips	FAIL
Telugu	FAIL	DiarBench 200, Kathbath 200, FLEURS 200	32 MB: 2 better; phone pack: 2 better	-1.20 (-2.21 to -0.41), 32 MB, DiarBench, English in Latin	Pack and Python differ on 2 of 1,679 clips	FAIL
Gujarati	MIXED	DiarBench 200, Kathbath 200, FLEURS 200	32 MB: 1 better	-0.16 (-0.28 to -0.04), 32 MB, LibriSpeech, English	Pack and Python differ on 2 of 1,679 clips	FAIL
Odia	FAIL	DiarBench 196, Kathbath 200, FLEURS 200	32 MB: 1 better; 53 MB: 1 better + 2 worse; phone pack: 2 worse	+3.29 (+0.58 to +6.46), 53 MB, DiarBench, English in Latin	Pack and Python differ on 2 of 1,675 clips; English letter share differs by over 1 point; 3 characters missing	FAIL
Kannada	FAIL	DiarBench 200, Kathbath 200, FLEURS 200	32 MB: 1 worse; 53 MB: 1 worse	+1.20 (+0.53 to +1.97), 53 MB, FLEURS, native	Pack and Python differ on 3 of 1,679 clips	FAIL
Urdu	MIXED	DiarBench 74, Kathbath 200, FLEURS 200	None	-	Pack and Python differ on 1 of 1,553 clips	MIXED

LANGUAGE	VERDICT	SETS (CLIPS)	ACCURACY MISSES AGAINST THE ENGINE	LARGEST MISS, POINTS (95% INTERVAL)	OTHER RULES	LINUX X86
Malayalam	MIXED	DiarBench 200, Kathbath 200, FLEURS 200	53 MB: 1 better	-0.45 (-0.92 to -0.02), 53 MB, Svarah, English	Pack and Python differ on 5 of 1,679 clips	MIXED
Assamese	MIXED	FLEURS 200, DiarBench 84	phone pack: 1 better	-0.80 (-1.50 to -0.19), phone pack, Svarah, language detected	Pack and Python differ on 1 of 1,363 clips	MIXED
Sanskrit	MIXED	IndicVoices 200, DiarBench 55	32 MB: 1 better	-0.47 (-0.93 to -0.02), 32 MB, Svarah, English	Pack and Python differ on 1 of 1,334 clips	MIXED
Punjabi	FAIL	DiarBench 200, Kathbath 200, FLEURS 200	32 MB: 1 better; phone pack: 1 better	-0.61 (-1.14 to -0.15), phone pack, FLEURS, native	Pack and Python differ on 2 of 1,679 clips	FAIL
Bodo	MIXED	IndicVoices 200, DiarBench 112	53 MB: 1 worse	+1.44 (+0.18 to +3.07), 53 MB, DiarBench, English in Latin	Pack and Python differ on 1 of 1,391 clips	FAIL
Dogri	FAIL	IndicVoices 200, DiarBench 24	32 MB: 2 worse; 53 MB: 1 better	+1.88 (+0.40 to +3.29), 32 MB, DiarBench, native	Pack and Python differ on 2 of 1,303 clips	FAIL
Kashmiri	MIXED	IndicVoices 200, DiarBench 32	None	-	Pack and Python differ on 2 of 1,311 clips	MIXED
Konkani	FAIL	IndicVoices 200, DiarBench 90	32 MB: 2 better; 53 MB: 2 better; phone pack: 1 better	-0.92 (-1.73 to -0.27), 53 MB, IndicVoices, English in Latin	Pack and Python differ on 1 of 1,369 clips	FAIL
Maithili	MIXED	IndicVoices 200, DiarBench 54	None	-	Pack and Python differ on 2 of 1,333 clips	FAIL
Manipuri	MIXED	IndicVoices 200, DiarBench 44	32 MB: 1 worse	+6.01 (+0.11 to +18.87), 32 MB, DiarBench, English in Latin	Pack and Python differ on 1 of 1,323 clips; English letter share differs by over 1 point	FAIL

LANGUAGE	VERDICT	SETS (CLIPS)	ACCURACY MISSES AGAINST THE ENGINE	LARGEST MISS, POINTS (95% INTERVAL)	OTHER RULES	LINUX X86
Nepali	MIXED	IndicVoices 200, DiarBench 45, FLEURS 200	phone pack: 1 worse	+1.45 (+0.21 to +2.94), phone pack, FLEURS, native	Pack and Python differ on 2 of 1,524 clips; English letter share differs by over 1 point	MIXED
Santali	MIXED	IndicVoices 200, DiarBench 99	32 MB: 1 worse	+1.79 (+0.04 to +4.33), 32 MB, DiarBench, English in Latin	Pack and Python differ on 2 of 1,378 clips	FAIL
Sindhi	FAIL	IndicVoices 200, DiarBench 27	53 MB: 2 better; phone pack: 2 better	-1.75 (-3.66 to -0.57), phone pack, DiarBench, English in Latin	Pack and Python differ on 2 of 1,306 clips	FAIL

Each miss in "Accuracy misses" is a paired interval that excludes 0, counted over rules 2 and 5. "Largest miss" is the difference from the engine in points of word error rate, so plus is worse. Set names are as in section 4.2. Every interval is in RESULTS.md, and the Linux column's in RESULTS_LINUX_X86.md.

A failed rule isn't always worse. The rules ask for a match, so a build can miss by scoring better as well as worse:

- **FAIL, with misses in the worse direction:** Hindi, Tamil, Odia, Kannada and Dogri. Odia is the clearest: its low-memory build is +3.29 on DiarBench, and it is the one language that lost characters (3).
- **FAIL, with accuracy misses only in the better direction:** Telugu, Punjabi, Konkani and Sindhi.
- **MIXED, with misses in the worse direction:** Marathi, Bengali, Bodo, Manipuri, Nepali and Santali. Manipuri's interval is wide, on 44 turns.
- **MIXED, with accuracy misses only in the better direction:** Gujarati, Malayalam, Assamese and Sanskrit.
- **No accuracy miss:** Urdu, Kashmiri and Maithili.

Across all 22 languages, 31 of 262 intervals under rule 2 missed, 16 worse and 15 better. If the builds matched the engine exactly, we estimate about 13 would miss by chance, though many intervals are not independent. Rules 1 and 6 held for every language, and the packs wrote no words on 100 zero and low-noise clips in any of them.

Two things we looked at after reading the rules. Neither changes a verdict:

- **Rule 3, every language missed it.** All but 3 of the 34 English mismatches between the pack and Python are on two English clips of 30 seconds or more, because Python cuts audio at 30 seconds and the pack does not. We confirmed that on one Urdu clip and estimate the others share the cause. Without those two clips, Urdu, Kashmiri and Maithili would miss no rule.
- **Hindi.** One DiarBench clip repeats a phrase in ONNX and not in the engine, and it holds 60 of the 80 extra errors. Without it, the native reference moves from +1.19 to +0.30 (0.00 to +0.61) on the 32 MB build and from +1.25 to +0.36 (+0.04 to +0.68) on the low-memory build.

10.3 Where the packs fall short

- **The engine is faster and lighter on a phone (Shruti Hindi, Galaxy F62):** a 10-second clip takes about 1.5 s including loading, in 63 MB, against about 2.4 s after a 3.4 s load, in about 190 MB for the sherpa-onnx pack. Across all 22 packs on that phone, 20 to 30 clips each, the real-time factor was 0.19 to 0.48, peak memory 254 to 387 MB and loading 3.7 to 10.4 s.
- **Repeats.** Now and then the model repeats a phrase: 117 of 9,536 test clips (1.2%) on the 4-bit build. The packages keep that output by default, as the engine and the pack do. Passing `loops="cut"` removes everything after the first copy, including any real words after it. One Sanskrit clip looped only on the phone, 140 words against 27 on the Mac, taking that language's phone error to 70.2 against 30.46 on 20 clips.
- **Long audio.** With the sherpa-onnx pack, split anything over 30 seconds with a voice-activity detector. Unsplit, a 60-second Hindi clip repeated a phrase after about 38 s. The Python and JavaScript packages cut audio into windows of at most 30 s wherever the cut falls.
- **Indian English.** The sherpa-onnx pack cannot set the language. On Hindi's Svarah set, word error was 29.4 with the language detected and 24.9 with English set. Across the 22 languages, the low-memory build took 51% to 78% of Svarah's clips for the model's own language, close to the engine each time.
- **Quiet.** The packs return nothing on digital silence and on the low-noise clips we tested. Real-world quiet needs a voice-activity detector.
- **Memory.** The 32 MB build uses about 0.8 to 1 GB after loading, because ONNX Runtime expands the weights, and the low-memory build about 120 MB (measured on Hindi, on a Mac).
- **Linux on x86: use the 32 MB build.** Its text was identical on a Mac and a Linux x86 machine (1,274 of 1,274 clips, Hindi and Urdu). The low-memory build and the phone pack were not. The verdicts above use the Mac for those two. Under the same rules, an AMD EPYC 7702P without VNNI gives 7 MIXED and 15 FAIL, with 48 misses under rule 2, and some gaps are large: +15.62 for Maithili on IndicVoices and +14.19 for Gujarati on DiarBench, on the low-

memory build. A likely cause, not yet tested, is 8-bit arithmetic on chips without VNNI. Chips with VNNI are untested.

- **Small differences between devices.** The same pack can differ by a letter or a word: on Hindi, 37 of 50 clips were identical between the phone and the Mac, and 38 of 50 between two sherpa-onnx library builds.
- **Output.** There is no punctuation, and numbers come out as words. Sindhi comes out in Devanagari.
- **Not tested.** An iPhone or the iOS Simulator, Windows, x86 chips with VNNI, phones other than the Galaxy F62, browsers other than one Chromium-based desktop browser, and the browser's microphone. Accuracy on other speakers and in noisy real use is not measured here.

10.4 The engine route: what you get

The engine route is the one section 6 measures. There is no key, no account and no server.

One `.cact` file per language: 29.5 MB for Shruti Hindi and 29.7 MB for the other 21, with 4-bit weights and 8-bit activations. The engine and the weights are both Apache-2.0. The engine is built for macOS, Linux, Windows, Android, iOS and WebAssembly; section 3 says which of these we measured.

10.5 The engine route: how to integrate

YOU ARE BUILDING	USE	WHERE
A Python app or a server	`pip install cactus-needle`, then `Whistle("shruti-tamil.cact")`	Section 3
A command-line tool	The `needle` binary, `needle --model shruti-tamil.cact --audio clip.wav`; add `--audio-stream` for live, word-by-word text	Section 3
A web page	The engine's WebAssembly build in a Web Worker, as in `examples/browser/`, which takes `?model=<language>`	Section 3
An Android or iOS app	Link the engine's C library (`needle.h`: `needle_load`, `needle_transcribe`, `needle_stream_transcribe_process`) and ship the model file with the app. Use 2 threads on phones with two fast cores	Engine's documentation

10.6 The engine route: setting the language

Leave the language on automatic for most uses, with two exceptions. Set `en` for speech that is entirely English, which the automatic choice often takes for the model's own language (section 8.1). Set `ne` when you know the speech is Nepali (section 8.2). In Python this is

`model.transcribe(audio, language="en")`, and on the command line `--audio-language en`. Section 3 lists every language code.

10.7 Where the engine route falls short

- **You link the engine yourself in a phone app.** We do not ship a ready-made Android or iOS library.
- **Not measured everywhere.** The 21 newer models have not been measured on Linux, Windows or iOS (section 3).
- **First use needs the network.** The model file downloads once unless you ship it with your app, and the Python package downloads Cactus's engine library on first use. After that it works offline.
- **Usage counts.** The `cactus-needle` Python package sends anonymous usage counts by default; set `NEEDLE_TELEMETRY=0` to turn that off. The command-line engine and the WebAssembly build make no network calls.
- **Output.** There is no punctuation, and numbers come out as words.

---

11. Release details

ITEM	WHERE
Weights and model cards	One repository per language: `udgamlabs/Shruti` for Hindi, `udgamlabs/shruti-<language>` for the others. Shruti collection
ONNX packs, with results for each language	udgamlabs/Shruti-ONNX; `pip install shruti-onnx`; `npm i shruti-onnx onnxruntime-web`
Demo (Hugging Face Space)	Shruti Space
Examples, docs and checksums	github.com/udgamlabs/shruti, with one results file per language under `docs/`
Model pages, with a try box that runs on your device	udgamlabs.com/models/shruti, one page per language
Shruti Hindi's own report	udgamlabs.com/research/shruti

Developer: Udgam Labs, udgamlabs.com. **Contact:** the Community tab of a model's Hugging Face page, or issues on the GitHub repository.

```
@techreport{udgamlabs2026shrutifamily,
  title       = {Shruti: 22 small offline speech-to-text models, one per
language},
  author      = {{Udgam Labs}},
  institution = {Udgam Labs},
  type        = {Technical report},
  year        = {2026},
  url         = {https://udgamlabs.com/research/shruti-family}
}
```

---

12. References

1. Cactus Compute. *Whistle* model card and weights. huggingface.co/Cactus-Compute/whistle. Engine: huggingface.co/Cactus-Compute/needle3.
2. Udgam Labs. *Shruti: offline speech-to-text for Hindi, English and Hinglish in 29.5 MB*, October 2026. udgamlabs.com/research/shruti.
3. Mehendale, D., Mehndiratta, A. et al. (2026). *Indic DiarBench: A Multilingual Joint Diarization and ASR Benchmark for Indian Languages.* arXiv:2607.23808. Dataset, CC BY 4.0, from Sarvam AI and AI4Bharat: sarvamai/indic-diarbench, revision 92877ba.
4. Conneau, A., Ma, M. et al. (2022). *FLEURS.* arXiv:2205.12446. Dataset: google/fleurs.
5. Javed, T., Nawale, J. A. et al. (2024). *IndicVoices.* arXiv:2403.01926. Dataset: ai4bharat/IndicVoices.
6. Javed, T., Bhogale, K. S. et al. (2022). *IndicSUPERB* (Kathbath). arXiv:2208.11761. Dataset: ai4bharat/Kathbath.
7. ARTPARK and IISc. *Vaani* transcribed speech. ARTPARK-IISc/Vaani-transcription-part.
8. SPRING Lab, IIT Madras. *SPRING-INX.* arXiv:2310.14654.
9. Panayotov, V. et al. (2015). *LibriSpeech.* ICASSP 2015. Dataset: openslr/librispeech_asr.
10. Ardila, R. et al. (2019). *Common Voice.* arXiv:1912.06670.
11. Radford, A. et al. (2022). *Whisper.* arXiv:2212.04356.
12. Sources for Shruti Hindi's sets (Monsoon, Lahaja, MUCS 2021, Svarah) and other models: section 13 of reference 2.
13. Each language's model card on Hugging Face and results file on GitHub.

About Udgam Labs. Udgam Labs is an independent AI research lab. We work from first principles: before we build, we understand why things work, whether we are training a model, fine-tuning one or applying AI to a new

problem. Udgam (उद्गम) is Sanskrit for origin. udgamlabs.com